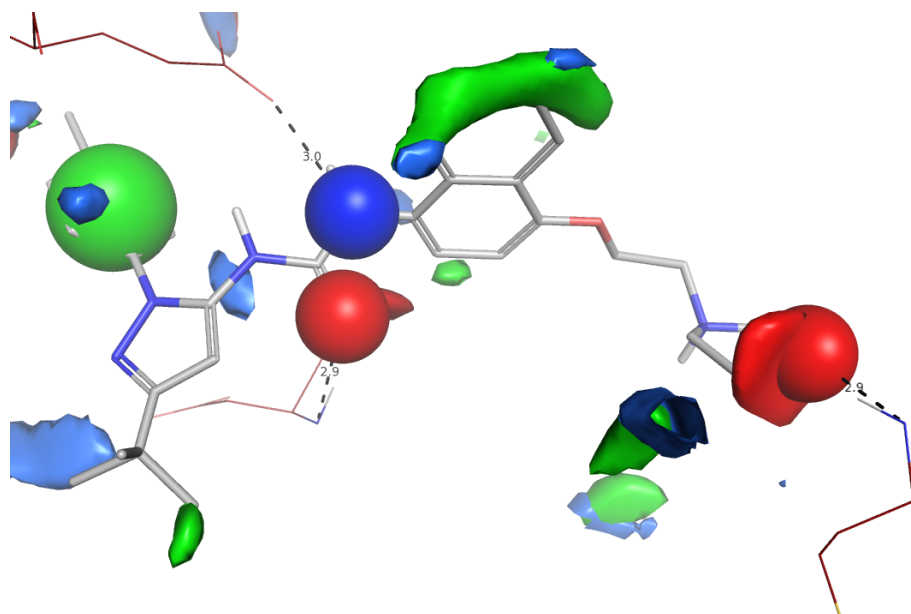




CavKA – Cavity Knowledge Acceleration

Entwicklung einer neuen strukturbasierten Methode zur Pharmakophor-Ableitung



Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Florian Kölling
aus Seesen

1. Referent:
2. Referent:
eingereicht am:
mündliche Prüfung (Disputation) am:

Professor Dr. Knut Baumann
Professor Dr. Conrad Kunick
04.06.2012
12.09.2012

Druckjahr 2012

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Vorträge auf Fachkongressen

(1) Cavity knowledge acceleration (CavKA) –
metamorphosis in automatic pharmacophore elucidation.

F. Koelling, K. Baumann

9th ICCS Meeting, Nordwijkerhout, The Netherlands (2011)

(2) Cavity knowledge acceleration (CavKA) –
from force fields to pharmacophores.

F. Koelling, K. Baumann

Doktorandenseminar, Braunschweig, Germany (2010)

(3) Phantastic ph4? –
three fully automatic pharmacophore elucidation methods under close scrutiny.

F. Koelling, K. Baumann

CCG User Group Meeting, Cambridge, UK (2010)

Posterbeiträge

(1) X-ploration space! Plif driven parallel screening (PLIDriPaS) in progress...

F. Koelling, K. Baumann

Summer School on Chemoinformatics, Strasbourg, France (2012)

(2) Cavity knowledge acceleration (CavKA) - on the ROCS!

F. Koelling, K. Baumann

OpenEye Cup, Santa Fe, New Mexico, USA (2012)

(3) CavKA^{HYBRID} – between hardspheres and gaussians...

F. Koelling, K. Baumann

German Conference on Chemoinformatics, Goslar, Germany (2011)

(4) CavKA – a new automatic pharmacophore elucidation method in progress...

F. Koelling, K. Baumann

German Conference on Chemoinformatics, Goslar, Germany (2010)

(5) Three steps ahead? –

a comparison of structure-based and ligand-based virtual screening methods.

F. Koelling, K. Baumann

DPHG Jahrestagung, Braunschweig, Germany (2010)

(6) Mr. Brightsite –

detecting hotspots ...and what we know when we only know little...

F. Koelling, K. Baumann

Sheffield Conference on Chemoinformatics, Sheffield, UK (2010)

(7) Detecting hotspots in binding sites –

validation of new fairplay approaches using MFSS and GRID.

F. Koelling, K. Baumann

DPHG Jahrestagung, Jena, Germany (2009)

*„Der wahre Weg geht über ein Seil,
das nicht in der Höhe gespannt ist,
sondern knapp über dem Boden.
Es scheint mehr bestimmt,
stolpern zu machen,
als begangen zu werden.“*

FRANZ KAFKA

Meinen Eltern.

Danksagungen

Die vorliegende Arbeit entstand in der Zeit von Juli 2007 bis Mai 2012 am Institut für Medizinische und Pharmazeutische Chemie der Technischen Universität Carolo-Wilhelmina zu Braunschweig unter der Betreuung von

PROFESSOR DR. KNUT BAUMANN.

Lieber Knut, ich möchte mich an dieser Stelle ganz herzlich für Deine Betreuung in den letzten Jahren sowie die gegebene Möglichkeit bedanken, mein Thema so selbstständig bearbeiten zu können. Danke für die vielen wertvollen Diskussionen auf freundschaftlicher Ebene und die zahlreichen Tagungen, an denen Du mir die Teilnahme ermöglicht hast und auf denen ich meine Arbeiten präsentieren durfte. Danke für Deine wertvolle konstruktive Kritik in der Endphase und für die Unterstützung bei all den anfallenden Aufgaben in meinem neuen Tätigkeitsfeld.

Mein besonderer Dank gilt Herrn

PROFESSOR DR. CONRAD KUNICK

der sich freundlicherweise dazu bereit erklärte, die arbeitsintensive Funktion des Zweitgutachters zu übernehmen.

Danke der nimmer müden SABRINA-VEGAS-WOLLENHAUPT für das Korrekturlesen dieser Arbeit und für ihre unglaublich wertvolle konstruktive Kritik. Danke für einen unvergesslichen Road-Trip in die 'verbotene Stadt' und zu 'langweiligen Erdlöchern'. Ich freue mich auf die kommenden gemeinsamen Projekte und auf den nächsten OE Cup (hoffentlich wieder ein Lufthansa Flug).

Danke an DR. JAN DREHER (guter Mann! - Du siehst ICH habe die GRIDs mit 0.5Å gerechnet ...). Du hast mir gezeigt, wie man die Python zählt und mir bei so vielen Problemen (insbesondere den programmiertechnischen) mit Rat und Tat zur Seite gestanden. Ein dickes Danke dafür - und für großartige Tagungsbesuche.

Mein tiefer und besonderer Dank gilt DR. MARKUS KOSSNER, der mich mit seiner entspannten Haltung und mit seinem enormen Wissen aus den Tälern der tiefen Frustration zu Beginn meiner Promotionszeit geführt hat. Danke für alles - und eine unvergessene Woche Cambridge!

Danke an Shantheya Balasupramaniam für ihren Support im Endspurt für meinen ersten wichtigen Vortrag.

Des Weiteren möchte ich all meinen gegenwärtigen sowie ehemaligen Kollegen des Arbeitskreises für ihre Zusammenarbeit und das nette Arbeitsumfeld danken.

Mein weiterer Dank gilt DR. LUTZ PREU für die intensive Unterstützung und Beratung bei der Durchführung meiner neuen Aufgabe(n) sowie der wertvollen Erfahrung, die ich im Rahmen meiner Assistententätigkeit im achten Semester sammeln durfte.

Danke an mein Lieblings-Semester (FIESEL, ANNA, SVENJA, JANINA, VERO, BORSTEL, ROBERT...). Ihr seid gute Freunde geworden, und ich verbinde mit Euch die Sternstunden meiner Assistentenzeit sowie viele wunderschöne Joker-Abende. Ihr habt einen wesentlichen Beitrag dazu geleistet, dass aus dem 'beliebtesten Assistenten von hinten' der 'beliebteste von vorne' geworden ist. Eure Abwesenheit wird hier schmerzlich vermisst ...

Danke meinen Mentoren ELKE und BERTRAM SCHÜTZE für ein wundervoll lehrreiches und persönlichkeitsbildendes praktisches (halbes) Jahr - eine für mich einschneidende und sehr wichtige Zeit, an die ich oft und gerne zurückdenke.

Danke an meine Eltern, die mich während meines Studiums, meines praktischen Jahres und meiner Promotion fortwährend unterstützt haben. Ihr habt immer an mich geglaubt und habt mich zum Weitermachen animiert - insbesondere dann, wenn es bei mir nicht gut lief. Ich möchte Euch von Herzen dafür danken!

Mein ganz besonderer und tiefer Dank gilt meiner Freundin INES KIELTSCH. Danke für Deine unermüdliche Unterstützung in all den (teilweise sehr schwierigen) Jahren. Danke, dass Du immer für mich da warst und da bist - ohne Dich würde ich hier heute nicht stehen.

Inhaltsverzeichnis

I	Einleitung	20
1	Screening im Wirkstoffdesign	21
1.1	Die Suche nach neuen Leitstrukturen	21
1.2	Experimentelles High Throughput Screening (HTS)	22
1.2.1	Zelluläre Wachstums- und Proliferations-Assays	23
1.2.2	Reporter-Gen-Assays	24
1.2.3	Enzym-Assays	24
1.3	Virtuelles In-Silico Screening (VS)	24
1.4	Methoden des Virtuellen Screenings	26
1.4.1	Ligandbasiertes Virtuelles Screening (LBVS)	26
1.4.1.1	1D-Methoden	27
1.4.1.2	2D-Methoden	27
1.4.1.3	3D-Methoden	27
1.4.2	Strukturbasiertes Virtuelles Screening (SBVS)	29
1.5	Scaffold Hopping	31
2	3D Pharmakophore	33
2.1	Definitionen und Einführung	33
2.2	Ligandbasierte 3D-Pharmakophore	37
2.3	Strukturbasierte 3D-Pharmakophore	37
3	Ligand-Rezeptor-Interaktionen	40
3.1	Chemisches Gleichgewicht und energetische Aspekte	40
3.2	Hydrophobe Interaktionen	43
3.3	H-Brücken-Donor- und -Akzeptor-Interaktionen	45

3.4	Ionische Interaktionen	46
3.5	Aromatische π -Interaktionen	47
4	Hotspot-Detektion in Bindetaschen	48
4.1	Wissensbasierte Ansätze	48
4.1.1	Isostar und Superstar	49
4.1.2	Relibase	49
4.1.3	Drugscore	49
4.1.4	Cavbase	49
4.2	Gitterbasierte Ansätze: GRID	50
4.3	Fragmentbasierte Ansätze: Multi Copy Simultaneous Search	52
5	GRID-Hotspots im strukturbasierten Wirkstoffdesign	54
5.1	Erstellung strukturbasierter Pharmakophore	55
5.2	Leitstruktur Optimierung	56
6	Methoden zur strukturbasierten Pharmakophor-Ableitung	57
6.1	Komplexbasierte Methoden: Geometrische Kriterien	58
6.1.1	Ligandscout	59
6.2	Komplexbasierte Methoden: Energetische Kriterien	60
6.2.1	E-Pharmacophores (Phase)	60
6.2.2	GBPM (GRID Based Pharmacophore Model)	60
6.3	Komplexbasierte Methoden: Geometrische und energetische Kriterien	61
6.3.1	Pocket 2.0	61
6.4	Methoden für Apo-Strukturen	62
6.4.1	Mr. Brightsite	63
6.4.2	LUDI (de-novo Design)	65
7	Ligandzentrische Ansätze	66
7.1	ROCS	68
7.2	EON	68
7.3	Piramid	69
7.4	Pharao	69
7.5	ShaEP	69
7.6	FLAP	70

8	Flexibilität von Datenbank-Molekülen	72
9	Motivation und Zielsetzung	74
II	Methoden	78
10	CavKA: Implementierung pharmakophorer Eigenschaften	79
10.1	Pharmakophore Ligandeigenschaften	79
10.1.1	Aromatische Eigenschaften	80
10.1.2	Hydrophobe Eigenschaften	80
10.1.3	H-Brücken-Donor-Eigenschaften	82
10.1.4	H-Brücken-Akzeptor-Eigenschaften	82
10.1.5	Ionische Eigenschaften	83
10.2	Pharmakophore Proteineigenschaften	83
10.2.1	Listen pharmakophorer Proteineigenschaften	87
10.2.2	Wörterbücher pharmakophorer Proteineigenschaften	88
10.2.3	Verkettete Listen aromatischer Proteineigenschaften	89
10.3	Pharmakophore Modelleigenschaften	89
10.3.1	MOE-Pharmakophore (.ph4)	91
10.3.2	Pharao-Pharmakophore (.phar)	91
10.3.3	Verbotene Sphären	91
11	CavKA: Implementierung der geometrischen Ligand-Rezeptor-Komplex-Analyse	92
11.1	Kompatible Ligand-Protein Eigenschafts-Kombinationen	92
11.2	CavKA: Hydrophobe Interaktionen	94
11.3	CavKA: H-Brücken-Interaktionen	95
11.4	CavKA: Ionische Interaktionen / Akzeptor-Metall-Interaktionen . . .	96
11.5	CavKA: Aromatische Interaktionen	96
11.5.1	Kationen- π -Interaktionen	96
11.5.2	'Face-to-Face'- $\pi - \pi$ -Interaktion	97
11.5.3	'Edge-to-Face'- $\pi - \pi$ -Interaktionen	99

12 CavKA: Implementierung der energetischen Bindetaschen-Analyse	101
12.1 Auswahl der Hotspot-Methode	101
12.2 GRID-Sonden	104
12.3 MIF-Berechnung	106
12.4 MIF-Adaption an das Ligand-Areal	106
12.5 MIF-Skalierung	107
12.6 MIF-Konturierung	108
 13 Die CavKA-Software	 110
13.1 Manuelle Zusammenführung geometrischer und energetischer Kriterien	110
13.2 MIF-basierte E-Pharmakophore	112
13.3 CavKA ^{HYBRID}	113
 14 CavKA: Validierung	 114
14.1 Modellvalidierung und Leistungskennzahlen	114
14.1.1 Validierung von Harte-Sphären-Modellen	116
14.1.2 Validierung von Gauß-Modellen	118
14.2 Generelle Probleme bei der Validierung strukturbasierter Pharmakophor- Modelle	122
14.2.1 Überschießende Anzahl pharmakophorer Sphären	122
14.2.2 Ein Komplex - ein Bindungsmodus	123
14.2.3 Verbotene Sphären	123
14.2.4 Methodenabhängigkeit der Suchergebnisse	124
14.3 Datensätze	124
14.3.1 Directory of Useful Decoys (DUD)	125
14.3.2 Maximum Unbiased Validation (MUV)	127
14.3.3 Fieldscreen-Datensatz	128
14.4 Daten-Vorbehandlungen	131
14.4.1 Zusammenstellung des Fieldscreen-Datensatzes	131
14.4.2 Vorbehandlung der Aktiven und Köder-Moleküle	132
14.4.2.1 Für Screenings mit strukturbasierten Pharmakophor- Modellen	132

14.4.2.2	Für Screenings mit ligandzentrischen Methoden . . .	132
14.5	Vorbehandlung der Screening Referenzen	134
14.5.1	Ligand-Rezeptor-Komplexe	134
14.5.2	Referenz-Liganden	135
14.6	Hitlistenprozessierung	135
15	PLIF Driven Parallel Screening (PLIDriPaS)	137
15.1	Protein Ligand Interaktions Fingerprint (PLIF)	137
15.2	Parallel Screening	140
15.3	PLIF Driven Parallel Screening (PLIDriPaS)	141
15.3.1	Datenvorbehandlung PLIDriPaS	141
15.3.2	Modellauswahl und Datenbanksuche	142
16	CavKA: Versuchsaufbau	144
16.1	Versuchsaufbau: Manuelle Zusammenführung geometrischer und ener- getischer Information	144
16.2	Versuchsaufbau: MIF-basierte E-Pharmakophore	145
16.3	Versuchsaufbau: CavKA ^{HYBRID}	146
17	Versuchsaufbau und Ziele: ligandzentrischer Ansätze	149
III	Ergebnisse und Diskussion	150
18	Ergebnisse und Diskussion: CavKA versus Ligandscout	151
18.1	CavKA: Vorteile durch MIF Information	154
18.1.1	CDK2: Identifikation hydrophober Interaktionen	154
18.1.2	AChE: Identifikation wichtiger Wassermoleküle	158
18.1.3	PDGFRB: Priorisierung pharmakophorer Eigenschaften	160
18.2	Einfluss der Referenz auf die Screening-Ergebnisse	162
18.2.1	ACE	162
18.2.2	FXa	164
18.2.3	P38	167
18.3	Einfluss der Konformeren-Auflösung	171

19 Ergebnisse und Diskussion: MIF-basierte E-Pharmakophore	173
20 Ergebnisse und Diskussion: PLIF Driven Parallel Screening (PLIDri-PaS)	176
20.1 PLIF-Berechnungen und alternative Bindungsmodi der PDE5	176
20.2 Auswahl der Referenzen für das ' <i>Parallel Screening</i> '	180
20.3 Ergebnisse PLIDriPaS (PDE5)	181
21 Ergebnisse und Diskussion: CavKA^{HYBRID}	183
21.1 Ergebnisse: Vergleich von CavKA ^{HYBRID} mit der ligandzentrischen Pharao-Methode	183
21.2 Diskussion: Vergleich von CavKA ^{HYBRID} mit der ligandzentrischen Pharao-Methode	185
21.2.1 Ligandzentrische Pharao-Methode	185
21.2.2 Verbotene Gauß-Sphären	185
21.2.3 Verbotene Harte-Sphären	185
22 Ergebnisse und Diskussion: Ligandzentrische Methoden im Ver- gleich	187
22.1 Ergebnisse ligandzentrischer Methoden: Einzelne Zielstrukturen . . .	187
22.1.1 ACE	188
22.1.2 AChE	189
22.1.3 CDK2	190
22.1.4 COX2	191
22.1.5 EGFR	192
22.1.6 FXa	193
22.1.7 HIV-RT	194
22.1.8 INHA	195
22.1.9 P38	196
22.1.10 PDE5	197
22.1.11 PDGFRB	198
22.1.12 SRC	199
22.1.13 VEGFR2	200
22.2 Ergebnisse ligandzentrischer Methoden: Methodenvergleich	201

22.3 Diskussion ligandzentrischer Methoden	205
22.3.1 Datensatzkritik und Massen Bias	205
22.3.2 Leistungsfähigkeit der ligandzentrischen Methoden	206
22.3.3 Abhängigkeit der Screening-Ergebnisse von der Referenz	209
23 Zusammenfassung	210
23.1 Zusammenfassung CavKA	210
23.2 Zusammenfassung ligandzentrische Methoden	211
24 Ausblick	213
 IV Anhang	 214
25 Module und Programme	215
25.1 Open-Source-Module	215
25.1.1 Numpy	215
25.1.2 hCluster	215
25.1.3 Biopython	216
25.1.4 Openbabel und Pybel	216
25.1.5 Pharao	216
25.1.6 Matplotlib und Pylab	216
25.2 Eigene CavKA-Programme	216
25.2.1 cavka.py	217
25.2.2 prepare_cavka_grid.py	218
25.3 Eigene CavKA-Module	218
25.3.1 sift_cavka_pharao.py	218
25.3.2 cavka_E.py	218
25.4 Weitere kommerzielle, bzw. frei verfügbare Programme	219
25.4.1 PyMOL	219
25.4.2 MOE	219
25.4.3 GRID	219
 26 Quellcode	 220
26.1 Modul sift_cavka.py	220

26.2	Modul cavka_E.py	234
26.3	Programm prepare_grid_cavka.py	235
26.4	Programm cavka.py	243
27	Parameter und Konfigurationsdateien	249
27.1	omega_HTS.parm (OMEGA)	250
27.2	omega_HQS.parm (OMEGA)	251
27.3	tautomers.param (QUACPAC)	252
27.4	pkatyper.param (QUACPAC)	252
27.5	GRID Parameter	253
27.5.1	grin.in	253
27.5.2	grid.in	254
28	Daten CavKA: Harte-Sphären-Modelle	255
28.1	Vergleich von CavKA mit Ligandscout	255
28.2	Einfluss der Konformeren-Auflösung	257
28.3	Leistungsfähigkeit von E-Pharmakophoren im Vergleich	258
29	PLIDriPaS	259
30	Daten: CavKA^{HYBRID}	260
30.1	CavKA ^{HYBRID} : ROC-Kurven	260
30.2	CavKA ^{HYBRID} : ROC-Anreicherungs-faktoren	264
31	Daten: Ligandzentrische Ansätze	265
31.1	Einfluss unterschiedlicher ROCS-Superpositionierungen auf die Screening-Ergebnisse mit EON .	265
31.2	Alternative P38-Referenz-Liganden	269
	Literaturverzeichnis	270

Abkürzungsverzeichnis

ACE Angiotensin Converting Enzym

AChE Acetylcholinesterase

ATP Adenosintriphosphat

AUC Area Under the Curve

BEDROC Boltzman-Enhanced Discrimination of ROC

CATS Chemically Advanced Template Search

CavKA Cavity Knowledge Acceleration

CDK Cyclinabhängige-Kinase

CDK2 Cyclinabhängige-Kinase 2

cGMP cyclisches Guanosin-5-Monophosphat

COX2 Cyclooxygenase 2

CSD Cambridge Structural Database

DHF Dihydrofolat

DHFR Dihydrofolat Reduktase

DUD Directory of Useful Decoys

EGFR Epidermal Growth Factor Receptor

FLAP Fingerprints for Ligands and Proteins

FXa Faktor Xa

GBPM Grid Based Pharmakophor Model

GMP Guanosin-5-Monophosphat

GPCR G-Protein gekoppelter Rezeptor

HIV-RT HIV-Reverse Transkriptase

HMG-Co-A-Reduktase 3-Hydroxy-3-Methylglutaryl-Coenzym-A-Reduktase

HTS High Throughput Screening

INHA Enoyl-ACP-Reductase

IUPAC International Union of Pure and Applied Chemistry

LBVS Ligand Based Virtual Screening

MCSS Multi Copy Simultaneous Search

MD Moleküldynamik

MIF Molekulares Interaktions Feld

MTX Methotrexat

MUV Maximum Unbiased Validation

P38 P38 Mitogen aktivierte Proteinkinase

PDB Protein Data Base

PDE5 Phosphodiesterase-5

PDGFRB Platelet Derived Growth Factor Receptor Kinase Beta

Pharao Pharmacophore Alignment and Optimization

PLIDriPaS PLIF Driven Parallel Screening

QSAR Quantitative Structure-Activity-Relationship

Relibase Receptor Ligand Database

RIE Robust Initial Enhancement

RMSD Root Mean Square Deviation

ROC Receiver-Operator-Characteristic

ROC-EF ROC-Anreicherungs-Faktor

ROCS Rapid Overlay of Chemical Structures

SAR Structure-Activity-Relationship

SASA Solvent Accessible Surface Area

SBVS Structure Based Virtual Screening

sc-PDB Screening PDB

ShaEP Shape and Electrostatic Potential

SRC Tyrosin Kinase SRC

TC Tanimoto Coefficient

VEGFR2 Vascular Endothelial Growth Factor Receptor 2

VS Virtuelles Screening

Teil I

Einleitung

Kapitel 1

Screening im Wirkstoffdesign

1.1 Die Suche nach neuen Leitstrukturen

Seit den Menschen bewusst ist, dass Krankheiten durch Substanzen geheilt oder Krankheitsverläufe gelindert werden können, sind sie auf der Suche nach neuen Molekülen, die jenes zu bewerkstelligen vermögen. Die Möglichkeiten solche Substanzen zu finden, sind sehr breit gefächert und reichen von der zufälligen Entdeckung, über die Isolierung von Naturstoffen aus Pflanzen, Pilzen, Mikroorganismen oder Tiergiften, bis hin zur Modifikation von bekannten körpereigenen Signalmolekülen. Die Isolierung und Aufreinigung, bzw. die Modifikation von Naturstoffen ist häufig problematisch, da genügend Ausgangsmaterial vorhanden sein muss und die Aufreinigung meist sehr aufwändig ist. Naturstoffe selbst sind häufig nur sehr schwierig synthetisch zugänglich¹.

Um neue Leitstrukturen zu finden, werden in der heutigen industriellen Praxis vor allem in der frühen Phase der Wirkstofffindung im Wesentlichen zwei Strategien verfolgt:

1. Das sogenannte *High Throughput Screening* (Abk. HTS): Beim HTS handelt es sich um einen *technologiebasierten* Ansatz², bei dem Substanzdatenbanken überwiegend automatisiert in einem Assay auf Bioaktivität getestet werden. Dieses Vorgehen wird auch als *random screening*³ bezeichnet und wurde mit der 'Suche nach der Nadel im Heuhaufen' verglichen⁴.
2. Das sogenannte *Virtuelle Screening* (Abk. VS): Beim VS handelt es sich um einen *wissensbasierten*² Ansatz. Auf Basis einer bereits bekannten Substanz mit biologischer Aktivität wird mit chemoinformatischen Methoden nach ähnlichen Substanzen gesucht, mit der Hoffnung, dass es sich bei diesen

Substanzen ebenfalls um biologisch aktive handelt. Anstatt zufällig mehrere Millionen möglicher Substanzen zu testen, wird dies nur mit beispielsweise den vielversprechendsten Einhundert getan. Im Gegensatz zum zufallsbasierten HTS stellt das VS einen rationalen Ansatz zur Wirkstofffindung dar.

Beide Ansätze werden mit ihren Vor- und Nachteilen nachfolgend kurz erläutert.

1.2 Experimentelles High Throughput Screening (HTS)

Im letzten Jahrzehnt war die Entwicklung der Biowissenschaften durch eine rasend schnelle Entwicklung und viele Innovationen geprägt. Die Sequenzierung des menschlichen Genoms, sowie die Sequenzierung der Genome vieler pathogener Krankheitserreger führte zu einer drastischen Zunahme neuer potentieller Zielstrukturen⁵.

Sobald eine Zielstruktur validiert worden ist, wird häufig mittels in-vitro Assays in firmeneigenen Substanzdatenbanken nach neuen Leitstrukturen gesucht. Die meist durch Robotertestsyste me automatisierten Arbeitsabläufe in einem HTS ermöglichen die Testung von 10000 bis 100000 Substanzen (engl. *compounds*) pro Tag⁵, bzw. von sogar mehr als 100000 Substanzen pro Tag⁵ (engl. *ultra high throughput screening*)⁶.

Üblicherweise werden die zu untersuchenden Substanzen in einem Primär-Assay (engl. *primary assay*) in einer Standard Konzentration (Einpunktmessung) auf biologische Aktivität getestet. Primäre Screenings sind jedoch häufig stark fehlerbehaftet⁷⁻⁹, weshalb positiv getestete Substanzen üblicherweise in einem Bestätigungs-Screen (engl. *confirmatory assay*) zusätzlich überprüft werden^{9,10}.¹ Substanzen, die biologische Aktivität in einem primären Screening aufweisen, werden häufig als *Hits* (quasi 'erste Treffer') bezeichnet. Die Inhibitionswirkung dieser *Hits* ist in den meisten Fällen noch nicht optimal und kann durch die Modifikation der Substanz immens gesteigert werden¹.

¹Problematisch sind aktive Substanzen, die im Primär-Screen als 'negativ' getestet worden sind und nicht erneut in einem Bestätigungs-Assay untersucht werden. Unter diesen Substanzen können sich trotzdem aktive Moleküle befinden (*falsch-negativ* Problematik)¹¹.

In Kombination mit den Entwicklungen in der kombinatorischen Chemie, die über automatisierte rekombinante Synthesewege den Zugang zu tausenden Substanzderivaten ermöglicht¹², ist das HTS ein wichtiger Ausgangspunkt für die Identifikation neuer Leitstrukturen. Im folgenden Abschnitt wird eine kurze Übersicht über häufig im HTS verwendete Assays gegeben. Abbildung 1.1 zeigt den schematischen Ablauf eines HTS.

Im Folgenden werden einige Assays mit ihren Vor- und Nachteilen kurz vorgestellt.

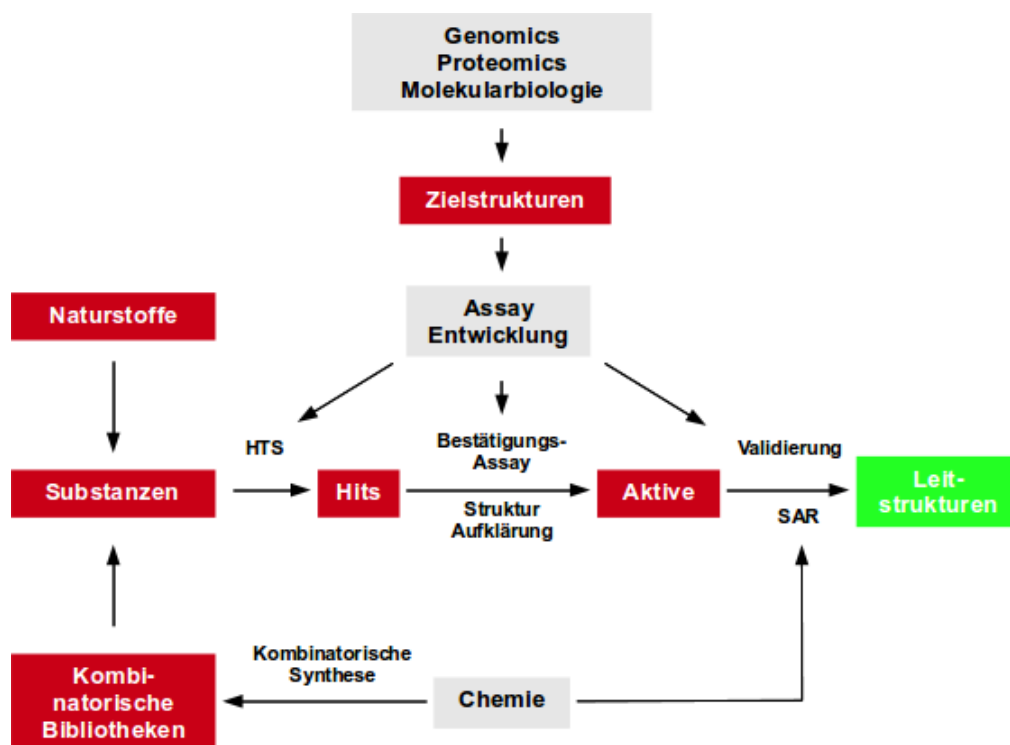


Abbildung 1.1: Schematischer Ablauf eines High-Throughput-Screenings (HTS). Abbildung aus Referenz¹³ (verändert nach Referenz¹⁴).

1.2.1 Zelluläre Wachstums- und Proliferations-Assays

Bei diesem Assay-Typus wird 'phänotypisch' analysiert, ob die zu untersuchende Substanz in der Lage ist, das Wachstum einer Zellkultur oder eines Gewebes¹⁴ zu inhibieren, weswegen diese Assays auch als 'phänotypische' Assays bezeichnet werden. Da nur das Wachstum einer Kultur evaluiert wird, ist es häufig sehr schwierig eine Aussage darüber zu treffen, ob es sich bei einer zu beobachtenden Wachstumsinhibition der Kultur um einen gewünschten pharmakologischen oder ungewünschten zytotoxischen Effekt handelt. Ferner kann keine Aussage darüber getroffen werden, welche Zielstruktur genau adressiert wird¹.

1.2.2 Reporter-Gen-Assays

Neue Fortschritte in der Molekularbiologie ermöglichten die Kultivierung von genetisch veränderten Zellen, die nach der Stimulation durch eine Substanz eine spezifische Reaktion liefern, beispielsweise die Expression eines Reporter-Gens, das über spektroskopische Methoden weiter analysiert wird (engl. *read-out*). Auch bei Reporter-Gen-Assays liegt keine isolierte Zielstruktur vor. Da nur der *read-out* ausgewertet wird, ist es denkbar, dass alternative Zielstrukturen an Stelle der eigentlich gewünschten Zielstruktur adressiert werden und die Expression des Reporter-Gens verhindert wird, bzw. dies über unspezifische Mechanismen geschieht¹.

1.2.3 Enzym-Assays

Mittlerweile ist es möglich Proteine rekombinant zu exprimieren und aufzureinigen. So ergibt sich die Möglichkeit eine Testung spezifisch an einem isolierten Enzym durchzuführen. Im Gegensatz zu den Ganzzell-Assays ist bei Enzym-Assays die Zielstruktur klar definiert, was jedoch nicht darüber hinwegtäuschen sollte, dass Assay-Artefakte, wie z.B. Aggregatbildung^{15,16}, unspezifisches Binden (engl. *non-leadlike binding*¹⁷, z.B. durch reaktive Gruppen¹⁸), optische Interferenz (*'the inner filter effect'*¹⁹) zu *falsch-positiven* Ergebnissen führen können.

1.3 Virtuelles In-Silico Screening (VS)

Virtuelle Screening-Methoden basieren auf dem Ähnlichkeitsprinzip²⁰⁻²³ (engl. *similarity property principle*), bei dem davon ausgegangen wird, dass 'ähnliche' Moleküle sich auch ähnlich bezüglich ihrer biologischen Aktivität verhalten²⁴. Während beim ligandbasierten Ansatz die Ähnlichkeit von zu vergleichenden Liganden bewertet wird, ist es beim strukturbasierten Ansatz die *sterische* und *chemische* Kompatibilität eines Liganden, die mit seiner Zielstruktur abgeglichen wird.

In beiden Fällen ist die Definition des zugrunde liegenden Ähnlichkeitsbegriffs alles andere als trivial: Allein die Fülle an Distanz- und Ähnlichkeitsmaßen auf der einen Seite und die Unmenge an mathematischen Repräsentationsmöglichkeiten der zu vergleichenden Moleküle über sogenannte *Molekül-Deskriptoren* auf der anderen Seite führen dazu, dass jede Methode 'Ähnlichkeit' unterschiedlich definieren kann. Im Vorfeld ist häufig nicht ersichtlich, welche Definition einen Datensatz ideal repräsentieren würde.

Ferner ist zu beachten, dass das Ähnlichkeitsprinzip nur begrenzt Gültigkeit besitzt: Bereits minimale Änderungen innerhalb eines Moleküls können große Änderungen in der Bioaktivität verursachen²². Man spricht in diesem Fall auch von Aktivitäts-Klippen (engl. *activity cliffs*)^{25,26}.

Auch wenn ein Virtuelles Screening das Experiment nie ersetzen wird (und auch nicht ersetzen soll), können sich beide Ansätze wertvoll ergänzen. Oft sind die Treffer-Raten eines HTS sehr niedrig und die Kosten immens hoch²⁷⁻²⁹. Trotz immenser Fortschritte bezüglich der experimentellen Durchführung im HTS hat dies nicht zu einer erheblich gesteigerten Markteinführung neuer Substanzen geführt¹².

Im Gegensatz hierzu steht der schier unendliche chemische Raum an theoretisch synthetisierbaren Substanzen: WALTERS et al.³⁰ schätzten die Anzahl potentieller Moleküle auf 10^{100} - was mehr wäre als die Anzahl an Atomen im Universum³⁰. Diese unglaubliche Fülle an denkbaren Molekülen lässt zwei Schlüsse zu¹³:

1. Es existiert mit sehr großer Wahrscheinlichkeit eine Substanz im chemischen Universum, die in der Lage ist den gewünschten biologischen Effekt an einer bestimmten Zielstruktur herbeizuführen.
2. Es ist unmöglich alle denkbaren Moleküle zu testen³⁰, bzw. nicht sehr wahrscheinlich durch ein zufälliges Screening neue Leitstrukturen zu finden: Substanzen (in diesem Falle Wirkstoffe), die in der Lage sind eine Zielstruktur spezifisch zu inhibieren, sind rar und ungleichmäßig im chemischen Raum verteilt. Meist liegen sie geballt in sogenannten *Clustern* vor³¹.

Eine sich hieraus ergebende Hauptanwendung des VS ist die Reduktion großer Datenbanken auf die Moleküle, die mit sehr hoher Wahrscheinlichkeit biologisch aktiv sind und mit vorhandenen Assay-Kapazitäten getestet werden können (engl. *focussed screening* oder auch *cherry picking*)³. Akademische Gruppen verwendeten VS bevorzugt, um in Substanzkatalogen kommerzieller Anbieter nach neuen Treffern zu suchen. VS ist jedoch keinesfalls hierauf beschränkt: Es können 'virtuelle', kombinatorisch erstellte Datenbanken durchsucht werden, deren Substanzen noch nicht einmal synthetisiert worden sind³².

KUBINYI gibt in einem Übersichtsartikel einen Überblick über Projekte, in denen Virtuelles Screening maßgeblich zum Erfolg beigetragen hat³³. Trotz beeindruckender Erfolge des VS wird von vielen Unternehmen auf ein fokussiertes Screening verzichtet und sämtliche verfügbare Substanzen in einem *in-vitro* Assay getestet^{3,34}. Dennoch werden HTS und VS sehr häufig parallel genutzt, jedoch in erster Linie um HTS Ergebnisse besser interpretieren zu können und die Qualität der Hits zu verbessern, anstatt Kosten zu senken³⁴. Möglichkeiten und Grenzen des Virtuellen Screenings werden von KLEBE in einem aktuellen Übersichtsartikel diskutiert³⁵.

1.4 Methoden des Virtuellen Screenings

Virtuelles Screening wird grob in ligand- und strukturbasierte Ansätze eingeteilt. Bei einem ligandbasierten Virtuellen Screening (engl. *ligand based virtual screening*; Abk. LBVS) wird nur Ligandinformation, unabhängig von der Zielstruktur, berücksichtigt. Bevorzugt wird LBVS verwendet, wenn die zu untersuchenden Zielstrukturen schwierig kristallographisch zugänglich sind, wie es beispielsweise bei Ionenkanälen³⁶, Transporter-Proteinen oder G-Protein gekoppelten Rezeptoren (Abk. GPCRs)³⁷ der Fall ist.

Die Voraussetzung für ein strukturbasiertes Virtuelles Screening (engl. *structure based virtual screening*; Abk. SBVS) ist das Vorhandensein von *struktureller Information*, entweder in Form von experimentell ermittelten NMR-, bzw. Röntgenkristall-Strukturen oder vorhergesagten Rezeptor-Homologie-Modellen³⁵. Ligandzentrische Ansätze nehmen eine Zwischenstellung zwischen SBVS und LBVS ein: Als Referenz wird hierbei ein co-kristallisierter Ligand verwendet. Abbildung 1.2 zeigt einen schematischen Vergleich von SBVS (Docking), LBVS (für jeweils unterschiedliche Deskriptoren) und einem ligandzentrischen Ansatz am Beispiel der *Cyclooxygenase-2*-Inhibitoren Celecoxib und Rofecoxib.

1.4.1 Ligandbasiertes Virtuelles Screening (LBVS)

Soll in einem Virtuellen Screening die Ähnlichkeit zwischen Referenz- und Datenbank-Molekül ermittelt werden, müssen die Moleküle in einem für den Rechner lesbaren Format hinterlegt sein. Eine solche 'mathematische' Beschreibung von Molekülen wird über sogenannte *Molekül-Deskriptoren* realisiert, die häufig auch als *Fingerprints* bezeichnet werden.

Es existieren unzählige Möglichkeiten ein Molekül mathematisch zu codieren³⁸, weshalb in diesem Abschnitt nur ein allgemeiner Überblick über gängige Grundprinzipien gegeben werden kann.

1.4.1.1 1D-Methoden

Bei 1D-Deskriptoren werden physikalisch-chemische Eigenschaften, wie z.B. die molekulare Masse, die Anzahl an Atomen eines bestimmten Typs (engl. *atom count*), die Anzahl an rotierbaren Bindungen durch *einen numerischen Wert* beschrieben. In vielen Fällen lassen sich die zur Deskriptor-Berechnung erforderlichen Eigenschaften direkt aus der Summenformel ableiten. Dies trifft nicht auf Eigenschaften, wie logP (dem Oktanol-Wasser Verteilungskoeffizienten) oder die Oberfläche eines Moleküls zu, die ebenfalls zu den 1D-Deskriptoren gezählt werden. Moleküle über 1D-Deskriptoren zu vergleichen ist die einfachste Methode, weswegen sie häufig im Englischen auch als '*Simple*'- oder '*Dumb*'-Deskriptoren bezeichnet werden. Trotz ihrer Einfachheit wurden mit ihnen häufig bessere Ergebnisse erzielt als mit komplexeren Methoden. BENDER und GLENN empfahlen daraufhin, nicht nur die absolute Leistung einer Methode, aktive Moleküle anzureichern, zu bewerten, sondern dies stets in Relation zu einem einfachen Deskriptor zu tun³⁹.

1.4.1.2 2D-Methoden

Bei den 2D-Methoden wird der *molekulare Graph* eines Moleküls, also die Konnektivität der einzelnen Atome, berücksichtigt, wie z.B. bei CATS 2D (Abk. *Chemically Advanced Template Search*)⁴⁰. Kontrovers wird die Einordnung von Methoden diskutiert, die die Anzahl bestimmter Substrukturen eines Moleküls codieren, wie z.B. die häufig verwendeten *MACCS-Keys*⁴¹. Sie werden aufgrund der Unabhängigkeit der Konnektivität von einigen Gruppen zu den 1D-Deskriptoren, von anderen Gruppen wegen der abgeprüften Mikro-Topologie (die jeweiligen Substruktur) zu den 2D-Deskriptoren gezählt.

1.4.1.3 3D-Methoden

Bei dreidimensionalen LBVS-Methoden wird die *geometrische* Anordnung von Atomen eines Moleküls im Raum codiert. Eine verwendete Referenz muss als dreidimensionale Struktur vorliegen. Hierfür können experimentell ermittelte NMR- oder Röntgenkristallstrukturen sowie vorhergesagte Modelle, wie z.B. ein energetisch günstiges Konformer verwendet werden.

Die 3D-Methoden unterscheiden sich vor allem bezüglich des Abgleichs von Referenz- und Datenbank-Molekül:

- Bei 3D-Deskriptoren, die von der Superpositionierung abhängen, wird versucht die codierten Eigenschaften von Referenz- und Datenbank-Molekül durch eine Superpositionierung maximal zur Deckung zu bringen (vgl. ligandzentrische Ansätze).
- Bei 3D-Deskriptoren, die unabhängig von einer Superpositionierung sind (wie z.B. MaP⁴², DIP²¹³, MOLPRINT 3D⁴³, Cats 3D⁴⁴, FLAP⁴⁵), wird die räumlich-relationale Anordnung molekularer Eigenschaften in einem Vektor codiert (beispielsweise die paarweisen Distanzen zwischen zwei H-Brücken-Donoren). 3D-Deskriptoren, die unabhängig von einer Superpositionierung sind, sind *rotations-* und *translations-invariant* (die paarweise Distanz zwischen den H-Brücken-Donoren ändert sich nicht, wenn das Molekül gedreht oder verschoben wird). Der Vergleich von Referenz- und Datenbank-Molekül wird somit über einen Vergleich der Vektoren realisiert, was einen deutlich geringeren Rechenaufwand bedeutet, als die Superpositionierung zweier Moleküle zu optimieren.

Als Faustregel gilt: Je mehr Dimensionen ein Deskriptor besitzt, desto mehr Rechenzeit ist erforderlich. Dieser Umstand sollte jedoch nicht zu der Fehlannahme verleiten, dass mit aufwändigeren Deskriptoren auch die besseren Ergebnisse erzielt werden. Dies ist nicht zwangsläufig der Fall^{39,46,47}. Die Recheneffizienz von 1D-Methoden und 2D-Methoden ermöglicht die Prozessierung deutlich größerer Datenbanken, als dies mit einer 3D-Methode möglich wäre, weshalb 1D- und 2D-Methoden häufig zur Reduktion besonders großer Datenbanken im Virtuellen Screening verwendet werden.

1.4.2 Strukturbasiertes Virtuelles Screening (SBVS)

Docking ist eine klassische Methode des strukturbasierten Wirkstoffdesigns, bei der ein Datenbank-Molekül in einer Bindetasche platziert (engl. *posing*) wird. Über eine Gütefunktion, eine sogenannte Scoring-Funktion (engl. *scoring function*) werden die räumliche Passgenauigkeit (engl. *shape*) und das Interaktionspotential zwischen Ligand und Rezeptor bewertet^{II}. Docking kann entweder für die Vorhersage eines Bindungsmodus von einem Molekül oder für ein Virtuelles Screening verwendet werden. In einem Screening gilt die Annahme, dass aktive Moleküle, die kompatibler zur Bindetasche sind, bessere Scores erhalten⁴⁸ und mit höherer Wahrscheinlichkeit biologisch aktiv sind. Während das Problem des *posings* leichter zu lösen ist (die korrekten Posen werden im Rahmen der Analyse des Konformerenraumes berechnet)⁴⁹, gelten Scoring-Funktionen im Allgemeinen als die Achillesverse des Dockings⁵⁰. Bewertungs-Funktionen im Docking scheitern in vielen Fällen daran, dass die korrekte Pose zwar berechnet worden ist, diese jedoch zu schlecht bewertet wird. In einem Virtuellen Screening ist dies insofern problematisch, als dass aktive Moleküle in diesen Fällen nicht auf den vorderen Rängen der Hitliste aufgeführt sind⁵¹.

^{II}'Rezeptor' wird in dieser Arbeit als Überbegriff für alle denkbaren biologischen Zielstrukturen (Ionenkanäle, Enzyme, DNA etc.) verwendet. Ebenso wird das Wort 'Ligand' als Überbegriff für Wirkstoffe (Rezeptor-Liganden, Agonisten, Antagonisten, Inhibitoren etc.) verwendet.

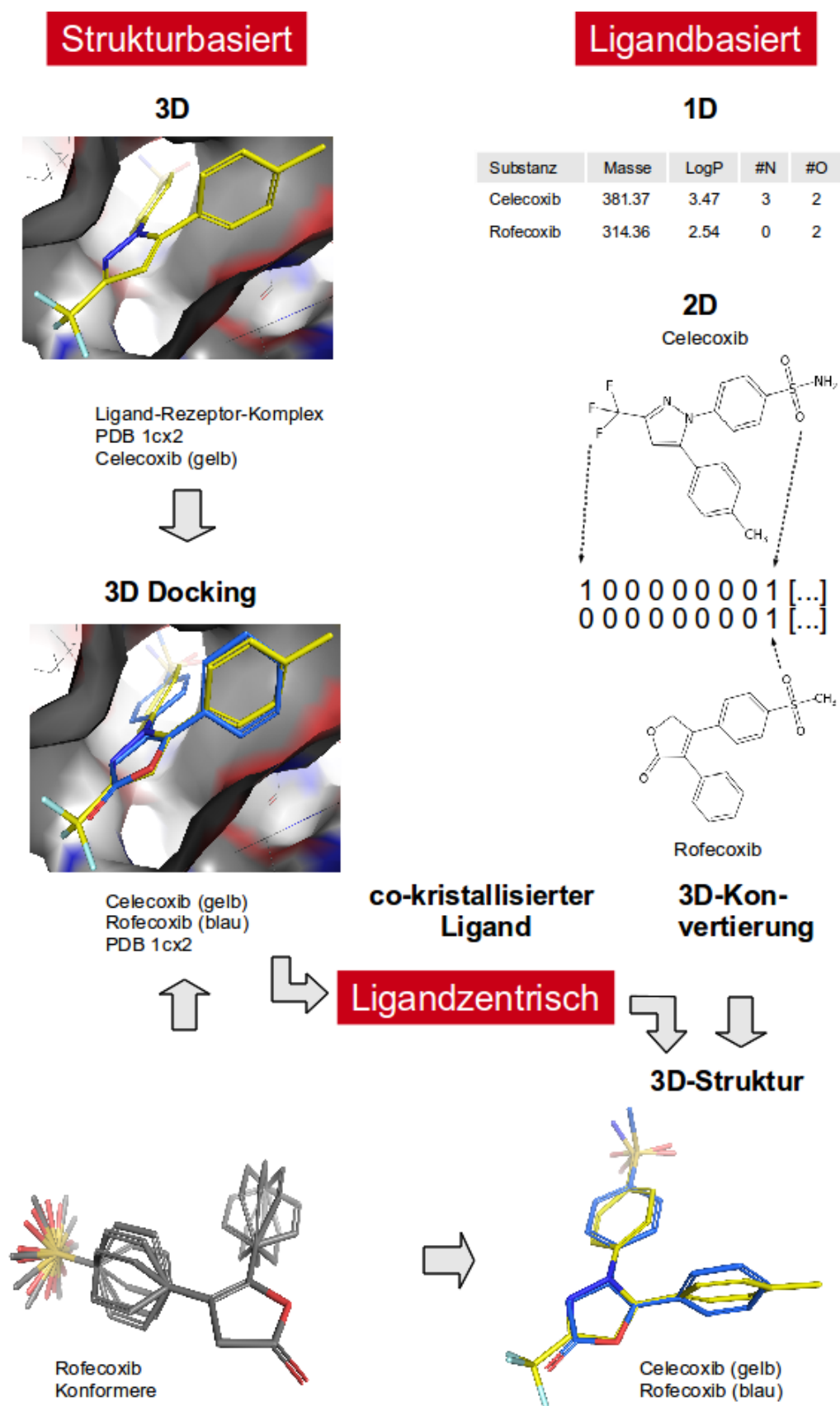


Abbildung 1.2: Schematische Übersicht struktur- und ligandbasierter sowie ligandzentrischer Ansätze im Virtuellen Screening.

1.5 Scaffold Hopping

Das Ziel eines Projektes in der Wirkstofffindung ist es neue, innovative Substanzen zu finden, die ein günstigeres pharmakodynamisches und / oder pharmakokinetisches Profil besitzen als die bereits Bekannten, weshalb häufig gezielt nach Strukturen mit diversen Grundgerüsten (engl. *scaffold*^{III}) gesucht wird. Auf Virtuelles Screening übertragen heißt dies, dass eine Methode im Idealfall Scaffold Hopping⁵²-Potential besitzt: Sie ist in der Lage, ausgehend von einer Referenz, biologisch aktive Moleküle anzureichern, die strukturell unterschiedlich und trotzdem 'isofunktional' zu der verwendeten Referenz sind⁵³.

Abbildung 1.3 zeigt ein sehr prominentes Beispiel für einen *Scaffold-Hopp* bei den bei Erektile Dysfunktion eingesetzten Phosphodiesterase-5 (Abk. PDE5)-Hemmern Sildenafil (Viagra®), Vardenafil (Levitra®) und Tadalafil (Cialis®)¹³. Sildenafil und Vardenafil unterscheiden sich nur minimal bezüglich der Stickstoffanordnung im Bipyridin sowie der Substitution an N4 des Piperazins (Methyl-Rest bei Sildenafil; Ethyl-Rest bei Vardenafil). Aufgrund dieser hohen Ähnlichkeit überrascht es nicht, dass die beiden Substanzen ein sehr ähnliches, pharmakodynamisches Profil aufweisen und die Halbwertszeiten mit je ca. 4h nahezu identisch sind⁵⁵. Tadalafil besitzt im Gegensatz dazu nicht nur ein völlig anderes Grundgerüst, sondern bindet darüber hinaus in einem anderen Bindungsmodus an seine Zielstruktur⁵⁶. Entsprechend unterschiedlich sind die pharmakokinetischen und pharmakodynamischen Parameter im Vergleich zu Sildenafil und Vardenafil^{57,58}: Tadalafil besitzt mit 17.5h eine deutlich längere Halbwertszeit sowie ein günstigeres Nebenwirkungsprofil⁵⁵.

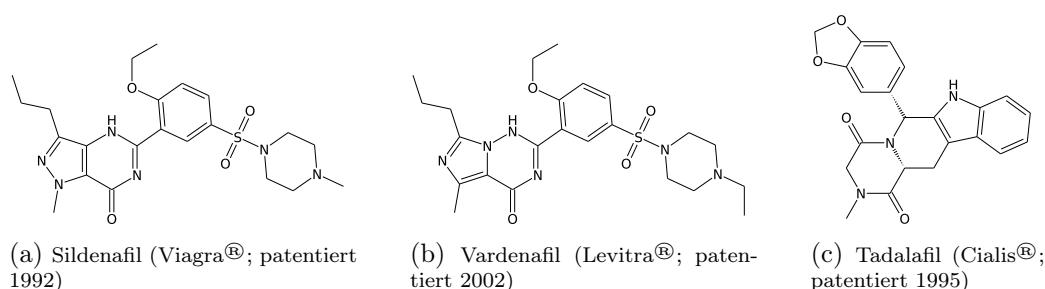


Abbildung 1.3: Strukturformeln von Sildenafil, Vardenafil und Tadalafil (Erläuterungen s. Text).

^{III}Für den Terminus 'Scaffold' existiert in der Literatur keine genaue und einhellig akzeptierte Definition. Er umschreibt üblicherweise das Grundgerüst eines Liganden⁵²⁻⁵⁴

Am Beispiel der oben genannten PDE5-Inhibitoren wird deutlich, dass in den meisten Fällen keine pharmazeutische Innovation zu erwarten ist, wenn Moleküle strukturell nur minimal modifiziert werden (s. Sildenafil und Vardenafil). Die Markteinführung von sogenannten *me-too* Präparaten wie Vardenafil zielt somit eher darauf ab, Marktanteile zu sichern und Patentrechte zu umgehen. Zusammenfassend können Vorteile durch *Scaffold Hopping* begründet sein durch (verändert nach BOEHM et al.⁵²):

- Veränderte pharmakodynamische Eigenschaften:
 - selektivere Substanzen (weniger unerwünschte Nebenwirkungen; s. Tadalafil)
 - alternative Bindungsmodi (Verbesserung der Resistenzlage; z.B. bei Antibiotika)
- Verbesserung pharmakokinetischer Eigenschaften:
 - verlängerte Halbwertszeiten (z.B. durch eine festere Bindung an die Zielstruktur)
 - verbesserte Löslichkeitseigenschaften (Ersetzen eines lipophilen Grundgerüsts durch einen hydrophileres)
 - Verbesserung der metabolischen Stabilität (Ersetzen metabolisch labiler Gruppen gegen stabilere)
 - Alternative Metabolisierung. Alternative Grundgerüste können somit ein unterschiedliches Wechselwirkungspotential besitzen.
 - Verbesserung der Resorption (z.B. der Austausch eines peptidischen Grundgerüsts gegen ein nicht-peptidisches)
- Patentrechtliche Gründe.

Im Folgenden Kapitel werden dreidimensionale Pharmakophor-Suchen als Technik im Virtuellen Screening vorgestellt. 3D-Pharmakophor-Modelle repräsentieren eine Screening-Referenz (entweder einen aktiven Liganden oder einen Ligand-Rezeptor-Komplex) in sehr abstrakter Art und Weise. Hieraus ergibt sich, dass jene Modelle meist ein recht hohes Scaffold Hopping-Potential besitzen⁴³.

Kapitel 2

3D Pharmakophore

2.1 Definitionen und Einführung

Bereits 1909 definierte PAUL EHRLICH das Wort Pharmakophor als molekulares Grundgerüst (gr.: *phoros*), das die *essentiellen* Eigenschaften eines Arzneistoffs (gr.: *pharmacon*) trägt, um biologische Aktivität zu gewährleisten⁵⁹. EHRLICHs Kernaussage spiegelt sich im Wesentlichen auch in der heute aktuellen IUPAC-Definition von 1998 wider:

“A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or to block) its biological response”⁶⁰.

In anderen Worten: Der Pharmakophor kann als eine Art Minimalvoraussetzung für Bioaktivität verstanden werden, die nicht auf spezielle, chemisch funktionelle Gruppen festgelegt ist. Anstatt dieser werden vielmehr die *Eigenschaften* jener Gruppen (engl. *features*) sowie deren räumliche Anordnung zueinander betrachtet.

Abbildung 2.1 zeigt schematisch das Vorgehen bei einer Pharmakophor Erstellung und einer anschließenden Datenbanksuche am Beispiel des *Cyclooxygenase-2* (Abk. COX2)-Inhibitors Celecoxib. Nahezu alle COX2-Inhibitoren besitzen einen zentralen, meist fünfgliedrigen aromatischen oder heteroaromatischen Ring mit zwei vicinalen Phenylringen, von denen einer eine Sulfon-Gruppe trägt⁶¹. Die einzelnen funktionellen Gruppen werden nun in pharmakophore Eigenschaften (Phenyl: aromatisch und hydrophob; Pyrazol: aromatisch; Sulfon-Sauerstoff: H-Brücken-Akzeptor) übersetzt (vgl. Abb. 2.1a).

Die Sphären stellen hierbei klar definierte Toleranzbereiche dar, in denen die jeweilige Eigenschaft anzutreffen sein muss (sog. Harte-Sphären-Modelle). Blendet man die Ligandinformation aus, so stellt die räumliche Anordnung der pharmakophoren Sphären zueinander eine sehr abstrakte Repräsentation des ursprünglichen Moleküls dar (vgl. Abb. 2.1b).

Solch ein Modell kann nun dazu genutzt werden, um in einer Multikonformeren-Datenbank nach Molekülen zu suchen, die den Pharmakophor der *Eigenschaft* nach (engl. *feature match*) und räumlich erfüllen (engl. *geometrical match*), um so neue Moleküle zu finden, die ebenfalls biologisch aktiv sind.

Im Gegensatz zu den ligandzentrischen Ansätzen wird bei den Harte-Sphären-Modellen binär klassifiziert, ob der Pharmakophor erfüllt (engl. *match*) oder nicht erfüllt ist¹. Eine Nicht-Erfüllung des Modells führt zum Verwerfen des Moleküls.

Der ebenfalls COX2-selektive Inhibitor Rofecoxib erfüllt das Pharmakophor-Modell *partiell* (engl. *partial match*): Alle Sphären werden räumlich und der Eigenschaft nach erfüllt bis auf das zentrale Lacton, das keine aromatische Eigenschaft besitzt (vgl. Abb. 2.1c). Pharmakophor-Modelle können also dahingehend modifiziert werden, dass nicht alle Sphären übereinstimmen müssen. So entstehen weniger restriktive Modelle. Auf der anderen Seite ist es auch möglich, Eigenschaften zu definieren, die obligat übereinstimmen müssen (engl. *must match* oder *essentials*). Dies ist vor allem dann sinnvoll, wenn Struktur-Wirkungsbeziehungen zu einer Zielstruktur bereits bekannt sind (wie beispielsweise die Interaktion zur sog. *Hinge*-Region bei den meisten Kinase-Inhibitoren).

Im Gegensatz zu Rofecoxib erfüllt der unselektive COX-Inhibitor Piroxicam⁶⁴ das zuvor definierte Pharmakophor-Modell weder räumlich noch den Eigenschaften nach (vgl. Abb. 2.1d). Piroxicam würde somit bei Verwendung des o.g. Modells in einer Datenbanksuche verworfen werden.

¹Manche Programme bieten zusätzlich die Möglichkeit, die Erfüllung eines Pharmakophor-Modells zusätzlich zu bewerten (z.B. über die Wurzel aus der Summe der mittleren quadratischen Abweichungen; engl. *root mean square deviation*; Abk. RMSD; wie z.B. in MOE⁶² oder einen Score wie z.B. in PHASE⁶³). Hierbei stellt sich jedoch grundsätzlich die Frage, ob ein Datenbank-Molekül vorzuziehen ist, das beispielsweise vier Eigenschaften des Pharmakophors mit einer sehr niedrigen RMSD erfüllt oder ein Molekül, das fünf Eigenschaften mit einer größeren RMSD erfüllt.

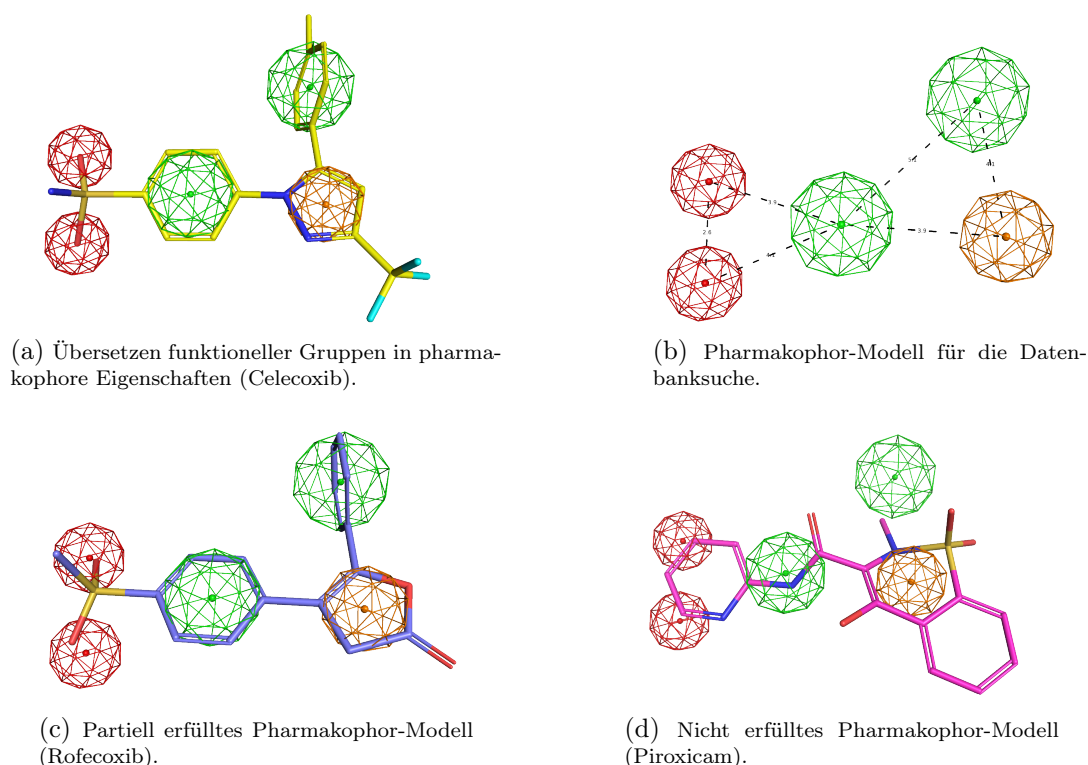


Abbildung 2.1: Schematische Darstellung einer Pharmakophor-Definition und Datenbanksuche am Beispiel von *Cyclooxygenase*-Inhibitoren. Farbcodierung pharmakophorer Ligandeigenschaften: grün: aromatisch und hydrophob; orange: aromatisch; rot: H-Brücken-Akzeptor.

Auch wenn eine Vielzahl von Methoden existiert, um Pharmakophor-Modelle auf Basis zweidimensionaler Information zu erstellen (z.B. CATS 2D^{40,65}; Feature Trees^{66,67}; für einen Übersichtsartikel über zweidimensionale Techniken siehe BROWN und JACOBY⁶⁸), wird in vielen Fällen bevorzugt mit dreidimensionalen Modellen gearbeitet. Dreidimensionale Modelle besitzen aufgrund ihrer abstrakteren Natur und höheren Unabhängigkeit vom Molekülgraphen häufig ein deutlich höheres *Scaffold Hopping-Potential*⁴³, was in zahlreichen Studien bestätigt worden ist^{69–72}. LANGER und KROVAT geben hierfür in einer Übersichtsarbeit zahlreiche Beispiele⁷³. Weitere Übersichtsarbeiten, die das *Scaffold Hopping-Potential* von Pharmakophor-Modellen thematisieren, sind kürzlich publiziert worden^{52,74}.

Es bleibt das Problem zu klären, *wie* Modelle erstellt, bzw. essentielle Schlüssel-Interaktionen identifiziert werden. Sind essentielle Eigenschaften bereits bekannt (z.B. durch eine Serie biologisch aktiver Moleküle oder einen bekannten Reaktionsmechanismus), können pharmakophore Sphären manuell definiert und platziert werden. Bei neuen Projekten, bei denen nur wenig über die Zielstruktur bekannt ist, ist dies meist nicht möglich.

Methoden der Chemieinformatik können beim automatisierten Erkennen mutmaßlich essentieller Eigenschaften sowie der Erstellung entsprechender Modelle (engl. *pharmacophore elucidation*) wertvolle Unterstützung bieten. So ist es möglich, Pharmakophor-Modelle auf der Basis mehrerer aktiver Liganden (s. Abb. 2.2a) zu erstellen (*ligandbasierter Ansatz*). Wenn kristallographische Daten verfügbar sind, können Pharmakophor-Modelle auf Basis eines Ligand-Rezeptor-Komplexes (s. Abb. 2.2b) oder einer Apo-Proteinstruktur (eine Protein-Struktur ohne co-kristallisierten Ligand) unter Verwendung von Kraftfeldern (s. Abb. 2.2c) abgeleitet werden (*strukturbasierte Ansätze*).

Im Folgenden wird eine kurze Übersicht über häufig verwendete Methoden zur Ableitung ligandbasierter und strukturbasierter Pharmakophor-Modelle gegeben. Gute Übersichtsarbeiten geben POPTODOROV et al.⁷⁵ und LEACH et al.⁷⁶. VAN DRIE thematisiert die historische Entwicklung von Pharmakophor-Modellen in einem weiteren Übersichtsartikel⁷⁷. Erfolgreiche Anwendungen von Pharmakophoren in der Wirkstofffindung werden von MANETTI et al. beschrieben⁷⁸.

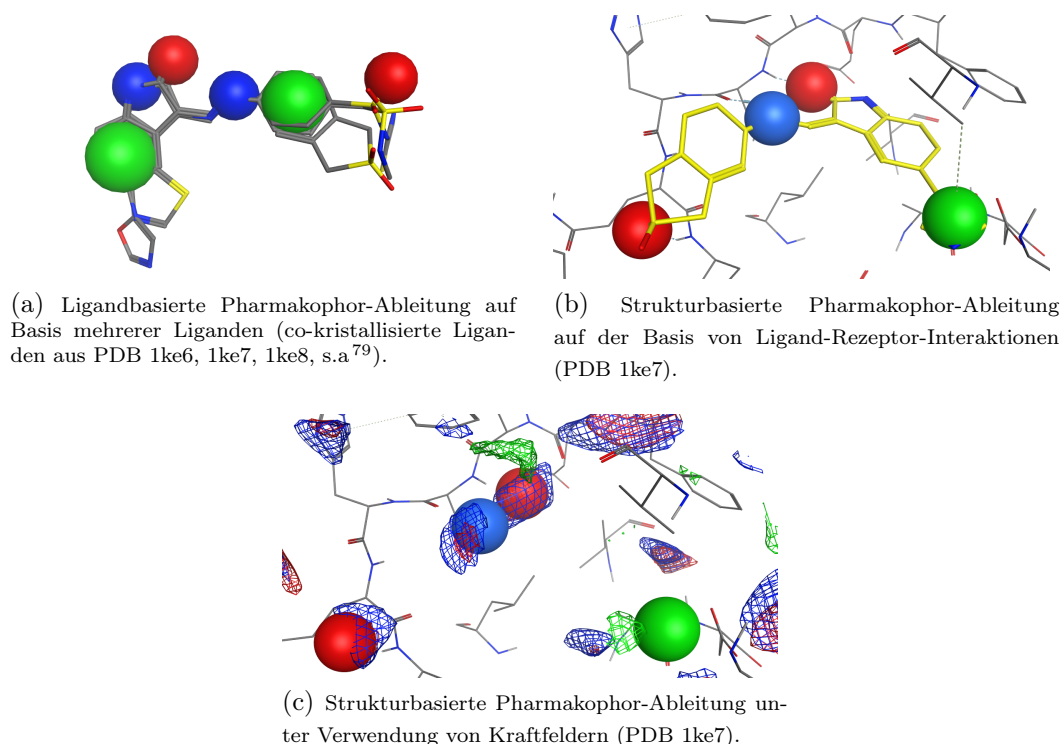


Abbildung 2.2: Unterschiedliche Ansätze der Pharmakophor-Ableitung am Beispiel der *Cyclinabhängigen-Kinase 2* (CDK2).

2.2 Ligandbasierte 3D-Pharmakophore

Bei einer ligandbasierten Pharmakophor-Ableitung werden Liganden mit biologischer Aktivität auf ihre räumliche Schnittmenge pharmakophorer Eigenschaften hin untersucht. Dazu werden die Liganden flexibel superpositioniert (engl. *flexible alignment*), d.h. zwei zu vergleichende Moleküle werden zunächst überlagert und ihre Überlappung in einem sich anschließenden Schritt maximiert (s.a. Abb. 2.2a). Je häufiger eine pharmakophore Ligandeigenschaft bei den verwendeten Referenz-Liganden nach der flexiblen Superpositionierung innerhalb eines Areals anzutreffen ist, desto höher wird die Relevanz der Ligandeigenschaft eingeschätzt. Ligandbasierte Pharmakophor-Ableitungen sind problematisch, wenn die verwendeten Referenz-Moleküle sehr unterschiedliche Grundgerüste besitzen und sich die Superpositionierung als schwierig gestaltet, bzw. wenn die Liganden in unterschiedlichen Bindungsmodi an ihre Zielstruktur binden.

In der jüngsten Vergangenheit sind viele verschiedene ligandbasierte Techniken zur Pharmakophor-Ableitung publiziert worden^{80–87}. Ferner bietet nahezu jede aktuelle Molecular Modeling Software die Möglichkeit ligandbasiert Pharmakophore abzuleiten (z.B. Phase⁶³, Ligandscout⁸⁸ und MOE⁶²).

2.3 Strukturbasierte 3D-Pharmakophore

Im Gegensatz zu ligandbasierten Pharmakophor-Modellen, die auf der Basis pharmakophorer Gemeinsamkeiten von aktiven Liganden erstellt werden, wird beim strukturbasierten Ansatz die zu untersuchende Zielstruktur (meist ein Protein) in Form einer Röntgenkristall-, einer NMR-Struktur oder eines Homologie-Modells analysiert. Auch wenn die Möglichkeit besteht strukturbasierte Pharmakophore auf der Basis von Apo-Proteinstrukturen zu erstellen (vgl. Abb. 2.2c), werden bevorzugt Ligand-Rezeptor-Komplexe zu diesem Zweck verwendet (vgl. Abb. 2.2b). Sie besitzen zum einen den Vorteil, dass der Ligand die Bindetasche bereits definiert, zum anderen sind durch die Ligand-Koordinaten Positionen von Eigenschaften eindeutig festgelegt, deren Lokalisierung mit Methoden zur *Hotspot*-Erkennung (s. Kap. 4) häufig nur bedingt möglich ist.

Damit ein Ligand biologisch aktiv ist, muss dieser gebunden an seiner Zielstruktur vorliegen. Dieses absolut fundamentale Kriterium, das die Grundlage des struktur-basierten Wirkstoff-Designs bildet, wurde bereits 1913 von PAUL EHRLICH mit dem lateinischen Satz:

“Corpora non agunt nisi fixata”⁸⁹

(frei übersetzt: „Die Medikamente tun nichts, wenn sie nicht gebunden werden“) formuliert. Eine Voraussetzung für eine solche Ligand-Rezeptor-Interaktion ist die sterische und elektronische Komplementarität von Ligand und Bindetasche, was EMIL FISCHER 1894 in seinem 'Schlüssel-Schloss'-Prinzip formulierte⁹⁰.

In einem Pharmakophor-Modell, das auf Basis von Ligand-Rezeptor-Komplexen erstellt wird, werden genau diese Aspekte der Komplementarität aufgegriffen - sowohl sterische als auch elektronische. Überträgt man EHRLICH'S Hypothese auf Pharmakophor-Modelle, so würde man Ligandeigenschaften, die an einer Rezeptor-Interaktion beteiligt sind (die den Liganden in seiner Bindetasche fixieren), eine größere Relevanz als den nicht-interagierenden Ligandeigenschaften einräumen^{II}. Zusätzlich zur Ligandinformation kann die Form der Bindetasche als ein weiteres Kriterium in einem Pharmakophor-Modell genutzt werden, da sie die Form des Schlosses definiert: Wünscht man, dass ein aktives Molekül in den dafür vorgesehenen pharmakophoren Sphären räumlich und der Eigenschaft nach liegt (engl. *included volumes*), so können die Atome der Bindetasche als verbotene Bereiche definiert werden, in die kein Datenbank-Molekül hineinragen darf (engl. *excluded volumes*). Die Verletzung einer verbotenen Sphäre durch ein Datenbank-Molekül führt zu dessen Ausschluss, was die Spezifität des Modells erhöht.

Auch beim Docking werden räumliche und chemische Komplementarität abgeglichen, und über eine Scoring-Funktion *quantitativ* bewertet. Dies erfordert deutlich mehr Rechenaufwand als bei Pharmakophor-Suchen, bei denen nur ein *qualitativer* Abgleich mit einem Pharmakophor-Modell stattfindet (Pharmakophor erfüllt, bzw. nicht erfüllt). Hierdurch ergibt sich ein klarer Vorteil für strukturbasierte Pharmakophore im Virtuellen Screening: Im Vergleich zum Docking sind Pharmakophor-Suchen erheblich schneller, eignen sich somit auch für große Datensätze und besitzen dennoch ein hohes *Scaffold Hopping-Potential*⁷⁶.

^{II}Im Umkehrschluss ist jedoch zu beachten, dass nur, weil es zu einer Interaktion zwischen Ligand und Bindetasche kommt, diese nicht zwingend notwendig für die biologische Aktivität sein muss⁹¹.

Zusammenfassend und auf FISCHERS Schlüssel-Schloss Prinzip übertragen kann der strukturbasierte Pharmokophor-Ansatz als Suche mit einem Schloss nach neuen, in das Schloss passenden Schlüsseln verstanden werden¹. Beim ligandbasierten Ansatz würde auf der Basis von Gemeinsamkeiten mehrerer passender Schlüssel nach wiederum neuen Schlüsseln gesucht werden.

Auch für strukturbasierte Ansätze bietet nahezu jede *Molecular Modeling Software* die Möglichkeit, Pharmakophor-Modelle zu erstellen und Datenbanksuchen durchzuführen. Als Auswahl seien MOE⁶², Phase⁶³, Catalyst⁹² und Ligandscout⁸⁸ genannt.

Ziel dieser Arbeit war die Entwicklung und die Validierung einer Methode zur strukturbasierten Ableitung von Pharmakophor-Modellen. Die im Rahmen dieser Arbeit entwickelte CavKA-Methode (Cavity Knowledge Acceleration) analysiert Ligand-Rezeptor-Komplexe automatisch nach geometrischen sowie energetischen Aspekten und erstellt auf dieser Basis Pharmakophor-Modelle. Die Kombination aus der geometrischen und energetischen Betrachtungsweise soll eine Identifikation relevanter Schlüsseleigenschaften (sogenannter *Hotspots*) ermöglichen. Berücksichtigt man diese in einem Pharmakophor-Modell, wird erwartet, aktive Moleküle in höherem Maße anzureichern als bei strukturbasierten Modellen, die nur anhand geometrischer Kriterien erstellt worden sind.

In Kapitel 3 werden daher zunächst allgemeine Grundlagen der Ligand-Rezeptor-Interaktion besprochen. In Kapitel 4 folgt eine allgemeine Darstellung von Methoden, die genutzt werden können, um *Hotspots* in einer Bindetasche zu finden. Kapitel 5 gibt eine Übersicht über erfolgreiche Anwendungsbeispiele, bei denen *Hotspots* gezielt im Wirkstoffdesign berücksichtigt worden sind, um potentere Inhibitoren zu entwickeln. Unterschiedliche, bereits publizierte Methoden, die strukturbasiert Pharmakophore ableiten, werden in Kapitel 6 vorgestellt. In Kapitel 7 werden zusätzlich verschiedene ligandzentrische Ansätze besprochen. Die Validierungsexperimente dieser ligandzentrischen Methoden, unter der Fragestellung mit welchen Methoden sich pharmakophore Eigenschaften idealerweise modellieren lassen, bilden den zweiten Schwerpunkt dieser Arbeit.

Kapitel 3

Ligand-Rezeptor-Interaktionen

Sollen auf Basis von Ligand-Rezeptor-Komplexen Pharmakophor-Modelle abgeleitet werden, werden diese in den meisten Fällen auf Interaktionen zwischen Ligand und Bindetasche hin untersucht. Eine zustande kommende Interaktion allein sagt jedoch nichts über deren Relevanz aus. Gelingt hingegen eine Einschätzung darüber, welchen Beitrag eine jeweilige Interaktion zur Bindungsenergie liefert, kann den interagierenden Ligandeigenschaften in einem Pharmakophor-Modell eine entsprechende Priorität eingeräumt werden.

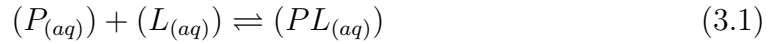
Im Folgenden wird ein allgemeiner Überblick über wichtige Ligand-Rezeptor-Interaktionen sowie deren energetische Aspekte gegeben. Jene bilden die theoretische Grundlage für die von CavKA vorgenommene Ligand-Rezeptor-Komplex-Analyse. Tabelle 3.1¹ fasst einige der in diesem Teil besprochenen Interaktionstypen zusammen und gibt einen qualitativen Überblick über deren Stärke¹.

Weitere, sehr umfangreiche Informationen über Ligand-Rezeptor-Interaktionen können in dem Buch *Protein-Ligand-Interactions*⁹³ gefunden werden, bzw. in einem exzellenten Übersichtsartikel⁹⁴ sowie der Dissertation von KOSSNER¹³. Die Implementierung der Interaktionserkennung von den hier vorgestellten Interaktionstypen durch CavKA wird detailliert in Teil II dieser Arbeit besprochen.

3.1 Chemisches Gleichgewicht und energetische Aspekte

Generell wird bei Ligand-Rezeptor-Interaktionen zwischen *kovalenten* und *nicht-kovalenten* Interaktionen unterschieden, wobei in dieser Arbeit und im Folgenden auf die *nicht-kovalenten* Interaktionen eingegangen wird.

Ligand-Rezeptor-Interaktionen finden im wässrigen Milieu statt, wobei der nicht gebundene Ligand ($L_{(aq)}$) und das freie, solvatisierte Protein ($P_{(aq)}$) im thermodynamischen Gleichgewicht mit dem assoziierten Ligand-Rezeptor-Komplex ($PL_{(aq)}$) stehen.



Aus den jeweils vorliegenden Aktivitäten von $(P_{(aq)})$, $(L_{(aq)})$ und $(PL_{(aq)})$ kann über das Massenwirkungsgesetz die Assoziationskonstante K_A berechnet werden:

$$K_A = \frac{c(PL_{(aq)})}{c(P_{(aq)}) \cdot c(L_{(aq)})} \quad (3.2)$$

K_A beschreibt somit die Stabilität des Ligand-Rezeptor-Komplexes. In der Praxis wird hingegen häufiger die Dissoziationskonstante K_D verwendet. Obwohl nicht gleich definiert, werden die Inhibitionskonstante K_i und K_D häufig synonym verwendet. K_D und K_i verhalten sich invers zu K_A und beschreiben, wie fest ein Ligand oder Inhibitor an seinen Komplex gebunden ist:

$$K_D = K_i = \frac{1}{K_A} = \frac{c(P_{(aq)}) \cdot c(L_{(aq)})}{c(PL_{(aq)})} \quad (3.3)$$

Die Werte für K_D , bzw. K_i werden in $\frac{\text{mol}}{\text{l}}$ angegeben. Experimentell ermittelte K_i -Werte bewegen sich zwischen $10^{-2} \frac{\text{mol}}{\text{l}}$ und $10^{-12} \frac{\text{mol}}{\text{l}}$ (vgl. BOEHM und KLEBE⁹⁵). Je kleiner die Werte, desto größer ist die Affinität des Liganden zum solvatisierten Protein, bzw. die Fähigkeit das Enzym zu inhibieren⁹⁴. Inhibitoren, die K_i -Werte im nanomolaren Bereich aufweisen ($10^{-9} \frac{\text{mol}}{\text{l}}$), können bereits als extrem potente Inhibitoren angesehen werden¹.

Es gilt die Annahme, dass Komplexbildung und Komplexzerfall in einem thermodynamischen Gleichgewicht stehen. Dies ermöglicht den Zusammenhang der energetischen Beziehungen von K_D , bzw. K_i aus der GIBBS Gleichung abzuleiten:

$$\Delta G^\circ = -R \cdot T \cdot \ln(K_i) \quad (3.4)$$

wobei ΔG° die freie Bindungsenthalpie (oder die freie GIBB'SCHE Energie), R die universelle Gas-Konstante (mit einem Wert von $8.314472 \frac{\text{J}}{\text{K} \cdot \text{mol}}$) und T die absolute Temperatur in Kelvin darstellt. Somit ist ΔG° die Änderung der freien Energie, die aus dem Gleichgewicht der Assoziation, bzw. Dissoziation von Ligand und Protein resultiert. Ist die Komplexbildung von Ligand und Rezeptor energetisch günstig und geht diese mit einem Gewinn der freien Energie einher, so wird ΔG° negativ.

Für die oben genannten K_i -Werte können somit ΔG° Werte von ca. $-10 \frac{\text{kJ}}{\text{mol}}$ bis $-70 \frac{\text{kJ}}{\text{mol}}$ (bzw. $2.3 \frac{\text{kcal}}{\text{mol}}$ bis $16.6 \frac{\text{kcal}}{\text{mol}}$)⁹⁵ erhalten werden.

Als Faustregel zum Abschätzen von K_D -Werten gilt: Eine Änderung des K_D -Wertes um eine Zehnerpotenz führt zu einer Änderung von $5.9 \frac{\text{kJ}}{\text{mol}}$, bzw. $1.5 \frac{\text{kcal}}{\text{mol}}$ der freien Energie¹.

Bevor die unterschiedlichen Interaktionstypen näher thematisiert werden, soll die *Freie Energie* ΔG° kurz erläutert werden. Diese setzt sich aus einem *enthalpischen* Term ΔH° und einem *entropischen* Term $-T \cdot \Delta S^\circ$ zusammen:

$$\Delta G^\circ = \Delta H^\circ - T \cdot \Delta S^\circ \quad (3.5)$$

Die Bindungsenthalpie ΔH° resultiert aus favorisierten Interaktionen zwischen Ligand und Bindetasche, wohingegen die Entropie ΔS° das Bestreben eines Systems beschreibt, in einen ungeordneteren Zustand überzugehen, bzw. seine inhärente Energie über möglichst viele Freiheitsgrade zu verteilen¹. ΔH° und ΔS° verhalten sich dabei gegenläufig: Bindet ein Ligand an seinen Rezeptor (Enthalpie steigt), so nimmt die Anzahl an Freiheitsgraden stark ab (die Translation und Rotation des Liganden ist stark eingeschränkt), weshalb die Ordnung des Systems zunimmt (Entropie sinkt). Hieraus folgt, dass sich Enthalpie und Entropie nicht isoliert voneinander betrachten lassen, sondern stets im Kontext analysiert werden müssen.

Tabelle 3.1: Experimentell, bzw. mit quantenmechanischen Berechnungen bestimmte paarweise Assoziationsenergien in der Gasphase (Tabelle aus *Wirkstoffdesign*¹). Die ermittelten Energien beziehen sich auf die Gasphase. Solvatations- und Desolvatationseffekte, wie sie bei Ligand-Rezeptor-Interaktionen eine entscheidende Rolle spielen, werden hierbei nicht berücksichtigt. Die Werte sollen allein einen qualitativen Überblick über die Stärke der jeweiligen Interaktion geben.

Dimer	Bindungsenergie in kJ/ mol
$\text{CH}_4 \cdots \text{CH}_4$	-2.0
$\text{C}_6\text{H}_6 \cdots \text{C}_6\text{H}_6$	-10
$\text{H}_2\text{O} \cdots \text{H}_2\text{O}$	-22
$\text{NH}_3 \cdots \text{NH}_3$	-18
$\text{Na}^+ \cdots \text{H}_2\text{O}$	-90
$\text{NH}_4^+ \cdots \text{CH}_3\text{COO}^-$	< -400

3.2 Hydrophobe Interaktionen

Obwohl hydrophobe Interaktionen im Vergleich zu ionischen oder polaren Wechselwirkungen in den meisten Fällen nominell nur relativ geringe enthalpische Beiträge zur freien Energie liefern (vgl. Tab. 3.1), handelt es sich um die wichtigste Kraft, die für die Stabilisierung von Biomembranen und Ligand-Rezeptor-Komplexen verantwortlich ist⁹⁶. Auch ΔG° -Werte allein lassen nur bedingt Rückschlüsse auf die Potenz eines Inhibitors zu. Beispielsweise sind die Ligand-Rezeptor-Interaktionen von Retinol und dem *Retinol bindenden Protein* ausschließlich hydrophober Natur: Obwohl für den Komplex nur ein ΔG° von $-38 \frac{\text{kJ}}{\text{mol}}$ ermittelt worden ist, beträgt der K_i -Wert beachtliche 190 nM⁹⁵.

Der eigentliche hydrophobe Effekt (engl. *'the hydrophobic effect'*), bzw. die Ursache, warum sich zwei lipophile Eigenschaften vermeintlich anziehen, ist keine direkte Wechselwirkung zwischen den Funktionalitäten selbst, sondern ist vielmehr indirekt auf die Verdrängung von Wassermolekülen zurückzuführen. Ein Wassermolekül besitzt mit seinen zwei Wasserstoffatomen zwei H-Brücken-Donoren und mit seinem Sauerstoffatom einen H-Brücken-Akzeptor, der erneut mit zwei H-Brücken-Donoren in Wechselwirkung treten kann. Hieraus folgt, dass Wassermoleküle über die H-Brücken-Bindungen untereinander ein regelmäßiges, geordnetes dreidimensional tetraedrisches Netzwerk von hoher Festigkeit ausbilden¹³ (vgl. die paarweisen Assoziationsenergien für Wasser in Tabelle 3.1).

Ein lipophiler Bindungspartner ist nicht in der Lage mit den Wassermolekülen über Wasserstoffbrücken in Wechselwirkung zu treten, was dazu führt, dass H-Brücken gebrochen werden und der regelmäßige Aufbau des Wassernetzwerkes gestört wird. Nach dem Bindungsbruch reorganisieren sich die Wassermoleküle jedoch um den Liganden, mit dem Bestreben die Anzahl der gebrochenen Bindungen zu minimieren und bilden mit dem Liganden ein Clathrat¹. Die den Liganden umgebenden Wassermoleküle sind in ihrer Beweglichkeit stark eingeschränkt, was sich entropisch ungünstig auswirkt. Bindet der solvatisierte Ligand an einen Rezeptor, so muss er zuvor seine Solvathülle abstreifen. Die bislang in ihrer Mobilität stark eingeschränkten Wassermoleküle können ins Bulkwasser diffundieren, was die Entropie des Systems erhöht⁹⁴. Ähnliches gilt für die solvatisierte Bindetasche: Gelingt es einem lipophilen Liganden Wassermoleküle aus hydrophoben Arealen der Bindetasche zu verdrängen, so erhöht dies die Entropie des Systems ebenfalls¹.

¹Clathrate oder Einschlussverbindungen (von lat.: *clatratus* = vergittert) sind chemische Komplexe, bei denen ein Gast-Molekül in mehrere Wirts-Moleküle (in diesem Fall Wassermoleküle) eingebettet ist. Die Wechselwirkungen zwischen Wirt und Gast sind hierbei *nicht-kovalenter* Natur.

In mehreren Studien wurde zwar ein Zusammenhang zwischen der lipophilen Kontaktoberfläche und der Freien Energie ermittelt^{97–102}, jedoch keine eindeutige Korrelation festgestellt. Der Versuch, die an der Ligand-Rezeptor-Interaktion beteiligte lipophile Kontaktoberfläche mit der experimentell ermittelten Bindungsenergie zu korrelieren, ergab schwankende Werte von -0.11 bis $-0.24 \frac{\text{kJ}}{\text{mol} \cdot \text{\AA}^2}$ als Beitrag zu ΔG° ^{103–105}. Basierend auf diesen Erkenntnissen formulierte BISSANTZ et al. als grobe Richtlinie, dass die perfekte Unterbringung einer Methyl-Gruppe (ca. 25\AA^3) in einer hydrophoben sub-Bindetasche zu einer Erhöhung der Bindungskonstante um den Faktor drei bis vier führt¹⁰⁶. Mangels eindeutiger Korrelation zwischen lipophiler Kontaktoberfläche und der Bindungskonstanten erschließt sich allenfalls ein *qualitativer* Zusammenhang⁹⁵. Auch wenn enthalpische und entropische Beiträge einer hydrophoben Interaktion zur freien Energie nur sehr schwierig vorherzusagen sind, so lassen sich dennoch die folgenden Schlüsse ziehen¹³:

1. Hydrophobe Wechselwirkungen sind der dominierende Faktor für die Ligand-Rezeptoraffinität^{1,106}.
2. Eine hydrophobe Interaktion findet nicht unmittelbar zwischen zwei lipophilen Gruppen statt. Sie ist vielmehr eine indirekte Folge der Verdrängung von Wassermolekülen. Daher lässt sich keine Direktionalität, wie beispielsweise bei H-Brücken (s. Abschn. 3.3) ausmachen. Bei den diffus erscheinenden hydrophoben Wechselwirkungen ist somit sehr schwer zu ermitteln,
 - (a) ob es überhaupt zu einer Interaktion zwischen der hydrophoben Ligand-eigenschaft und der Bindetasche kommt und
 - (b) welche der hydrophoben Ligandeigenschaften für die biologische Aktivität respektive die Erstellung eines Pharmakophor-Modells tatsächlich relevant sind⁷⁶.

Beide Fragen werden weiter im Rahmen der CavKA-Implementierung in Teil II erörtert.

3.3 H-Brücken-Donor- und -Akzeptor-Interaktionen

Interaktionen, die über H-Brücken vermittelt werden, zählen zu den elektrostatischen Interaktionen⁹⁴. Hierbei interagiert ein elektronegatives Atom wie Stickstoff, Sauerstoff oder Fluor (H-Brücken-Akzeptor) mit einem Wasserstoffatom, das seinerseits kovalent an ein ebenfalls elektronegatives Atom gebunden ist (H-Brücken-Donor). H-Brücken-Bindungen sind nicht-kovalenter Natur. Sie sind deutlich stärker als van-der-Waals Interaktionen jedoch schwächer als ionische oder kovalente Interaktionen.

Der Beitrag von ausgebildeten H-Brücken zur Bindungsenergie unterliegt großen Schwankungen und hängt stark vom jeweiligen Einzelfall ab^{93,95}. Als Konsequenz lässt sich nicht allein aufgrund ihrer Anzahl auf eine hohe Bindungskonstante schließen. So werden beispielsweise in den Ligand-Rezeptor Komplexen von *Glucose-Glykogen-Phosphorylase*¹⁰⁷, *Sialinsäure-Hämagglutinin*¹⁰⁸, *Glutathion-Glutathion-S-Transferase*¹⁰⁹ sehr viele H-Brücken-Bindungen und nahezu keine nennenswerten hydrophoben Ligandeigenschaften beobachtet. Die Bindungskonstanten sind trotz der vielen H-Brücken-Bindungen überraschend gering⁹⁴.

Da sowohl Ligand und Protein in solvatisierter Form vorliegen, spielen Solvatationseffekte eine entscheidende Rolle, die schlussendlich darüber entscheiden, ob durch das Ausbilden einer H-Brücke der Betrag an freier Energie zunimmt oder nicht. Da die Assoziation zwischen den Wassermolekülen selbst sehr groß ist (vgl. Tab. 3.1) muss sehr viel Energie aufgewandt werden, um diese Bindung zu brechen⁹⁵. So kann die Ausbildung einer zusätzlichen, jedoch energetisch ungünstigen H-Brücke, zu einem Verlust des Betrages der freien Energie führen¹¹⁰. Die Stärke einer H-Brückenbindung hängt von vielen Faktoren wie Solvatationseffekten, der lokalen Dielektrizitätskonstanten, Entropieeffekten uvm. ab. Diese sind nur sehr schwierig mit theoretischen Methoden vorherzusagen. So gelingt es häufig nur unbefriedigend energetische Beiträge einer H-Brückenbindung zur freien Energie *in-silico* vorherzusagen oder Bindungskonstanten mit der Anzahl von H-Brücken zu korrelieren^{94,95}. Als Faustregel kann davon ausgegangen werden, dass, wenn es einem Liganden gelingt mehr H-Brücken mit seiner Bindetasche auszubilden als dies vorher mit seiner Solvathülle möglich war, der Betrag an freier Energie zunimmt¹.

Auch wenn die Stärke der Interaktion bei H-Brücken-Bindungen schwer vorherzusagen ist, gelang es deren räumliche Geometrie über statistische Auswertungen von Röntgenkristallstrukturen recht genau zu definieren. MILLS und DEAN untersuchten als eine der Ersten H-Brücken-Bindungsgeometrien unterschiedlicher pharmazeutisch relevanter Substrukturen in verschiedenen Ligand-Rezeptor-Komplexen¹¹¹.

Aufgrund der stetig wachsenden Anzahl von Röntgenkristallstrukturen konnten Ansätze wie Isostar¹¹², Superstar¹¹³ (kleine Moleküle analysierend), bzw. Relibase^{114,115} und Cavbase¹¹⁶ (Proteine, bzw. Ligand-Rezeptor-Komplexe analysierend) erfolgreich genutzt werden, um Bindungsgeometrien zu beschreiben.

Aus den Erkenntnissen all dieser Studien lassen sich relativ einfache geometrische Regeln definieren, die einen Ligand-Rezeptor-Komplex auf eingegangene H-Brücken-Interaktionen hin untersuchen. Zusammenfassend lässt sich über die H-Brücken-Interaktionen festhalten¹³:

1. H-Brücken-Bindungen sind stark directional. Sie spielen bei der Ligand-Rezeptor-Erkennung die wichtigste Rolle^{95,106}.
2. Der Beitrag zur Bindungsenergie ist bei H-Brücken-Bindungen nur schwer abschätzbar, die räumliche Geometrie ist hingegen gut beschrieben^{94,95,106}.

3.4 Ionische Interaktionen

Ionische Interaktionen zählen wie H-Brücken-Bindungen zu den *elektrostatischen Interaktionen* und werden auch als Salzbrücken bezeichnet. Sie werden vorzugsweise zwischen positiv und negativ geladenen Funktionalitäten ausgebildet. Salzbrücken sind deutlich fester als eine H-Brückenbindung (s. Tab. 3.1) und deutlich seltener als diese. Aufgrund ihrer Stärke stellen sie häufig entscheidende Schlüssel-Interaktionen, wie z.B. bei Faktor Xa dar¹, weswegen ionischen Interaktionen stets eine besondere Aufmerksamkeit gelten sollte.

Für ionische Wechselwirkungen allein kann keine Direktionalität nach festen geometrischen Regeln definiert werden, sondern 'nur' eine Distanzabhängigkeit. Da jedoch ionische Interaktionen häufig vergesellschaftet mit H-Brücken-Bindungen auftreten, ist es oft schwierig, eine reine ionische Interaktion von der H-Brücken-Interaktion abzugrenzen, weswegen auch der Terminus *ladungsunterstützte Wasserstoffbrücke* verwendet wird¹³. Um ionische Interaktionen richtig vorhersagen zu können, ist die Definition korrekter Protonierungszustände eine unmittelbare Voraussetzung. Auch hier ist die jeweils lokal vorliegende Dielektrizitätskonstante verantwortlich für die pK_A -Werte saurer und basischer Gruppen in der Bindetasche und entscheidet letztlich über deren Protonierungszustand⁹⁴.

Auch wenn ionische Interaktionen meist einen sehr hohen Beitrag zur freien Energie liefern, sollte dies nicht darüber hinwegtäuschen, dass ebenfalls eine starke Abhängigkeit von Solvations- und Desolvations-Effekten besteht.

3.5 Aromatische π -Interaktionen

In jüngster Vergangenheit wurde die Rolle spezifischer aromatischer Interaktionen verstärkt diskutiert. Bereits 1985 entdeckten BURLEY et al. starke geometrische Abhängigkeiten aromatischer Interaktionen, die sich nicht über Hydrophobizität erklären ließen¹¹⁷. Die Arbeitsgruppe führte diese Beobachtung auf das starke Quadrupolmoment von π -Elektronen-Systemen zurück, was von anderen Arbeitsgruppen bestätigt worden ist¹¹⁸.

In einem aromatischen System stehen p -Orbitale senkrecht zur Ringebene, die wiederum mit benachbarten p -Orbitalen überlappen, was zu einem delokalisierten π -Elektronen-System führt. So zeichnen sich aromatische Systeme durch eine hohe Elektronendichte über- und unterhalb der Ringebene aus¹¹⁹, die analog zu H-Brücken-Akzeptoren in der Lage sind, mit positiv oder partiell positiv geladenen Spezies zu interagieren. COLE et. al. bestätigten diese Theorie empirisch über Kontaktstatistiken aromatischer Interaktionen¹²⁰ auf Basis von Isostar¹¹².

Über delokalisierte π -Elektronen vermittelte Ligand-Rezeptor-Interaktionen stellen einen wichtigen und - im Gegensatz zu den ungerichteten hydrophoben Interaktionen - einen gerichteten Interaktionstypus dar. π - π -Interaktionen können sich beispielsweise zwischen aromatischen Ligandeigenschaften und den Seitenketten der aromatischen Aminosäuren Phenylalanin, Tyrosin und Tryptophan ausbilden¹²¹. Hierbei wird zwischen einer sogenannten *Face-to-Face*-(die Ringebenen ordnen sich parallel zueinander an) und *Edge-to-Face*-(die Ringe ordnen sich senkrecht zueinander an) Anordnung unterschieden¹²².

Aufgrund der hohen Elektronendichte unter und über der Ringebene lassen sich auch bevorzugt Interaktionen mit Kationen beobachten⁹⁴. Solche Kationen- π -Interaktionen stellen beispielsweise in der Bindetasche der *Acetylcholinesterase* eine entscheidende Schlüssel-Interaktion dar^{123,124}. Sie zeigen darüber hinaus eine beachtliche Festigkeit: In einem direkten Vergleich von Salzbrücken und Kationen- π -Interaktionen in wässriger Lösung wurde für letztere ein um bis zu $-10 \frac{\text{kJ}}{\text{mol}}$ größerer Beitrag zur Bindungsaffinität ermittelt¹²⁵.

Die unterschiedlichen Typen der Ligand-Rezeptor-Interaktionen belegen eindrucksvoll deren Vielfältigkeit, bzw. wie schwierig energetische Beiträge zur freien Energie abzuschätzen sind. In Kapitel 11 und 12 werden die CavKA Implementierungen, wie Ligand-Rezeptor-Komplexe nach geometrischen und energetischen Kriterien untersucht werden, detailliert beschrieben. Ziel dieses kombinierten Ansatzes ist es, wichtige Schlüssel-Interaktionen zu identifizieren, um diese gezielt in einer Pharmakophor-Hypothese zu berücksichtigen.

Kapitel 4

Hotspot-Detektion in Bindetaschen

Nicht alle Bereiche in einer Bindetasche sind gleich wichtig. Areale, in denen ein Ligand bevorzugt mit der Bindetasche interagiert und ein besonders großer Beitrag zur Bindungsenergie aus einer solchen Ligand-Rezeptor-Interaktion resultiert, werden als *Hotspots* bezeichnet¹.

In Abwesenheit von co-kristallisierten Liganden ist es mitunter sehr schwierig solche *Hotspots* zu identifizieren, insbesondere dann, wenn wenig Information über die Zielstruktur vorhanden oder wenig über den Reaktionsmechanismus eines Enzyms bekannt ist. In der Vergangenheit sind daher große Anstrengungen unternommen worden, Werkzeuge zu entwickeln, die in der Lage sind solche *Hotspots* vorherzusagen. Die Entwicklung von Liganden, die gezielt an diesen Stellen der Bindetasche eine Ligand-Rezeptor-Interaktion eingehen, verspricht hochaktive Substanzen zu liefern (s. Kap. 5). Bei den Methoden zur *Hotspot*-Detektion wird grob zwischen wissensbasierten und kraftfeldbasierten Methoden unterschieden. Bei letzteren wird zusätzlich zwischen gitterbasierten und fragmentbasierten Ansätzen differenziert.

4.1 Wissensbasierte Ansätze

Bei wissensbasierten Ansätzen werden Interaktionen in Kristallstrukturen statistisch analysiert und auf dieser Basis Regeln für den jeweiligen Interaktionstypus definiert. So kann beispielsweise ein Carbonyl-Sauerstoff-Atom (H-Brücken-Akzeptor) als Suchanfrage verwendet werden, um zu prüfen innerhalb welcher Distanzen und Winkel es beispielsweise zu einer Interaktion mit einer Hydroxyl-Gruppe (H-

¹In der Literatur existiert keine einheitliche Definition für das Wort *Hotspots*. Üblicherweise werden mit diesem Terminus besonders wichtige und für biologische Aktivität essentielle Schlüssel-Stellen in der Protein-Bindetasche beschrieben.

Brücken-Donor) kommt. Ein häufiges Vorkommen einer Interaktion gilt hierbei als wahrscheinlicher und energetisch günstiger als eine selten vorkommende Interaktion. Diese Statistiken lassen sich zur *Hotspot*-Detektion nutzen.

4.1.1 Isostar und Superstar

Die Isostar-Datenbank¹¹² vom Cambridge Crystallographic Data Center (Abk. CSD) beinhaltet kleine Moleküle, die mit dem Programm Superstar¹²⁶ zur Anzeige und Konturierung von *Hotspots* in der Bindetasche verwendet werden können.

4.1.2 Relibase

Relibase ist ein Datenbanksystem, das gezielt verwendet werden kann, um nach bestimmten Ligand-Rezeptor-Interaktionen in Ligand-Rezeptor-Komplexen zu suchen^{114,115}. Relibase weist konzeptionell viele Analogien zu Isostar auf. Im Gegensatz zu Isostar, das sich auf kleine Moleküle beschränkt, bilden bei Relibase Ligand-Rezeptor-Komplexe die Grundlage für die statistische Analyse.

4.1.3 Drugscore

Der von GOHLKE et al. entwickelte Drugscore ist eine wissensbasierte Scoring-Funktion für die Bewertung von Ligand-Rezeptor-Interaktionen¹²⁷. Für die Parametrisierung wurde Relibase als Datengrundlage verwendet. Ligand-Rezeptor-Interaktionen wurden analysiert und Histogramme für die Verteilung bestimmter Kontakte zwischen Ligandeigenschaften und Aminosäuren der Bindetasche erstellt. Auch bei diesem Ansatz wurde häufiger zu beobachtenden Interaktionen eine größere Relevanz als den seltener zu Beobachtenden eingeräumt.

4.1.4 Cavbase

Das in Relibase integrierte Cavbase ist eine von SCHMITT et al. entwickelte Methode für den Vergleich von Protein-Bindetaschen¹¹⁶. Cavbase identifiziert anhand geometrischer Regeln Aminosäuren des Proteins, die aufgrund ihrer Ausrichtung und der Nähe zur Oberfläche in der Lage sind, ihre pharmakophoren Eigenschaften mit in die Bindetasche einzubringen (sogenannte *cavity flanking residues*). Sie werden als punktförmiges 'Pseudo-Zentrum' dargestellt.

Eine Bindetasche wird durch solch einen Satz Pseudo-Zentren repräsentiert. Mittels *Clique Detection*^{II} können Bindetaschen über die Pseudo-Zentren unabhängig von ihrer Sequenz miteinander verglichen werden.

4.2 Gitterbasierte Ansätze: GRID

Das bereits 1985 von PETER GOODFORD entwickelte Programm GRID war einer der ersten Ansätze, um *Hotspots* in Protein-Bindetaschen zu berechnen und zu visualisieren¹²⁹. Im Rahmen der Methode wird eine zu untersuchende Röntgenkristallstruktur in ein dreidimensionales Gitter eingebettet. Anschließend werden kugelförmige Sonden (engl. *probes*) mit unterschiedlichen pharmakophoren Eigenschaften (vgl. Tab. 4.1) auf jedem Gitterpunkt positioniert und ein Interaktionspotential zwischen der Sonde und dem Protein berechnet.

Das GRID-Potential E_{xyz} (in *kcal/mol*) an dem Gitterpunkt mit den Koordinaten_{xyz} ist hierbei die Summe eines Lennard-Jones (E_{LJ})-, eines Elektrostatik (E_{el})- und eines direktionalen H-Brücken-Potentials (E_{HB})^{130–132}:

$$E_{xyz} = \sum E_{LJ} + E_{el} + E_{HB} \quad (4.1)$$

In dem so erhaltenen Molekularen Interaktions-Feld (Abk. MIF) wird zwischen attraktiven Kräften ($E_{xyz} < 0$) und repulsiven Kräften ($E_{xyz} > 0$) unterschieden. Die GRID-Potentialfunktion ist empirisch für eine Vielzahl von Kristallstrukturen parametrisiert worden.

Die für eine Sonde berechneten MIFs können konturiert, d.h. entsprechend ihrer Interaktionsenergie dargestellt werden, sodass z.B. nur 20% der Punkte mit dem höchsten Interaktionspotential betrachtet werden. Bereiche mit einem sehr negativen Potential (also stark konturierte Felder) lassen darauf schließen, dass es sich hierbei um *Hotspots* handelt.

Die Anwendung GRIDs ist keinesfalls auf Protein-Bindetaschen festgelegt. Statt MIFs ausschließlich von Protein-Bindetaschen zu berechnen, können auch MIFs von Protein-Interfaces oder von Liganden berechnet werden. Die sich ergebenden Ein-

^{II}In der Graphentheorie (einem Teilgebiet der Mathematik) wird Graph durch sogenannte *Knoten* beschrieben, die über sogenannte *Kanten* verbunden sind. In der Chemieinformatik werden Moleküle häufig als Graphen interpretiert (Atome stellen hierbei die Knoten, Atom-Bindungen die Kanten dar). Sind alle Knoten des Graphen paarweise über Kanten miteinander verbunden (vollständiger Graph) spricht man von einer *Clique*. Cliques stellen in den meisten Fällen eine Untermenge eines Graphen dar. Algorithmen zur sogenannten *Clique Detection* (z.B. der BRONKERBOSCH Algorithmus¹²⁸) werden häufig zum Vergleich von Molekülen oder, wie im Falle von Cavbase, zum Vergleich von Bindetaschen verwendet.

satzgebiete sind vielfältig (wie z.B. Virtuelles Screening¹³³; Scaffold Hopping¹³⁴; die Analyse Quantitativer Struktur-Wirkungsbeziehungen, engl. *quantitative structure-activity-relationship*, Abk. QSAR¹³⁵; Vorhersagen pharmakokinetischer Eigenschaften^{136,137}, metabolischer Stabilität¹³⁸, von pK_A -Werten¹³⁹ und Tautomeren¹⁴⁰). Einen Überblick über die unterschiedlichen Anwendungen sowie noch detailliertere Informationen geben CROSS und CRUCIANI in einem Übersichtsartikel¹⁴¹, bzw. sind in dem Buch *Molecular Interaction Fields*¹⁴² zu finden. CROSS und CRUCIANI thematisieren in einem weiteren Übersichtsartikel die Verwendung GRIDs zur Ableitung strukturbasierter Pharmakophore¹⁴³.

Auch die Grundidee GRIDs ist von anderen Arbeitsgruppen aufgegriffen und analog implementiert worden (vgl. Score¹⁴⁴, MOE⁶² *Interaction Potential*). Aufgrund der besseren Visualisierungsmöglichkeiten in MOE ist in Abbildung 4.1 eine beispielhafte Konturierung eines Interaktionsfeldes dargestellt.

Tabelle 4.1: Pharmakophore Eigenschaften einiger GRID-Sonden.

Sonde	Eigenschaft
DRY	hydrophob
C3	hydrophob / sterisch
N1	H-Brücken-Donor
O	H-Brücken-Akzeptor
N1+	kationisch
O-	anionisch
H2O	H-Brücken-Donor / H-Brücken-Akzeptor (Wasser-Sonde)

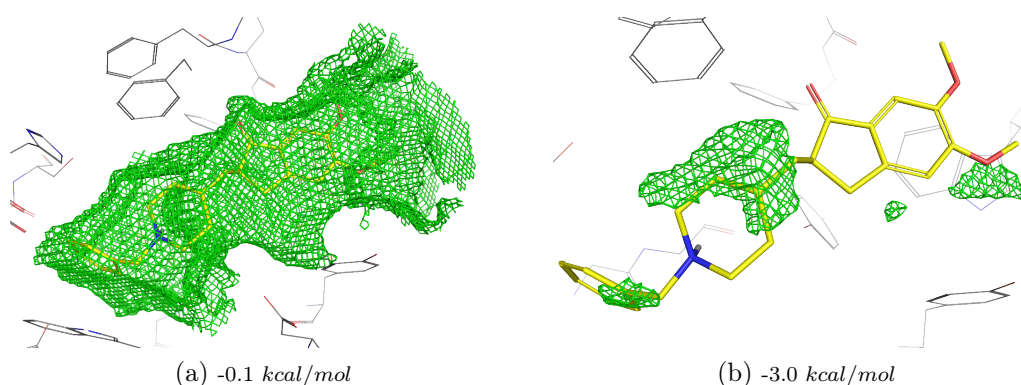


Abbildung 4.1: Konturierung von MOE *Interaction Potential*-Feldern der 'Dry'-Sonde (aromatisch-hydrophobe Eigenschaften). Die von der Bindetasche berechneten Felder sind mit Abweichungen in der Lage, die hydrophoben Areale in Abwesenheit von Ligandinformation (Ligand: Donepezil) vorherzusagen (PDB 1eve; *Acetylcholinesterase*).

4.3 Fragmentbasierte Ansätze: Multi Copy Simultaneous Search

Die Arbeitsgruppe von MARTIN KARPLUS stellte 1992 mit *Multi Copy Simultaneous Search* (Abk. MCSS) den ersten fragmentbasierten Ansatz zur *Hotspot*-Detektion vor^{145,146}. Statt das zu untersuchende Protein in ein Gitter einzubetten, werden kleine Fragmente mit unterschiedlichen pharmakophoren Eigenschaften (s. Tab. 4.2) zufällig in der Bindetasche platziert. Über ein Kraftfeld (CHARMM¹⁴⁷) werden die Fragmente simultan, ohne sich dabei gegenseitig zu beeinflussen, im Hinblick auf die Interaktionsenergie energieoptimiert und bewertet. Die Optimierung erfolgt entweder so lange, bis eine maximale Anzahl von Zyklen durchlaufen oder ein lokales Optimum für das entsprechende Fragment gefunden worden ist und sich der Score nicht mehr durch weitere Optimierung verbessern lässt. Indizieren zwei sehr dicht beieinander liegende Fragmente dasselbe lokale Minimum, so wird eines dieser Fragmente gelöscht und der nächste Optimierungszyklus mit dem verbliebenen Fragment durchlaufen. So nimmt die Zahl der initial platzierten Fragmente stetig ab. Die am Ende verbleibenden Fragmente verweisen in Verbindung mit ihrem Score auf potentielle *Hotspots*.

Das MCSS-Prinzip ist von mehreren Gruppen in ähnlicher Weise implementiert worden^{148,149}. SCHUBERT und STULTZ geben einen Überblick über die MCSS-Methode sowie erfolgreiche Anwendungen¹⁵⁰. In Abbildung 4.2 ist die in MOE⁶² implementierte *Multi Fragment Simultaneous Search* (Abk. MFSS) als Analogon zu MCSS verwendet worden, um die Methodik fragmentbasierter Ansätze exemplarisch darzustellen.

Tabelle 4.2: Pharmakophore Eigenschaften einiger MCSS-Fragmente.

Fragment	Eigenschaft
Ammonium	kationisch
Acetat	anionisch
Aceton	H-Brücken-Akzeptor
Dimethylamin	H-Brücken-Donor
Methan	hydrophob
Benzol	hydrophob-aromatisch

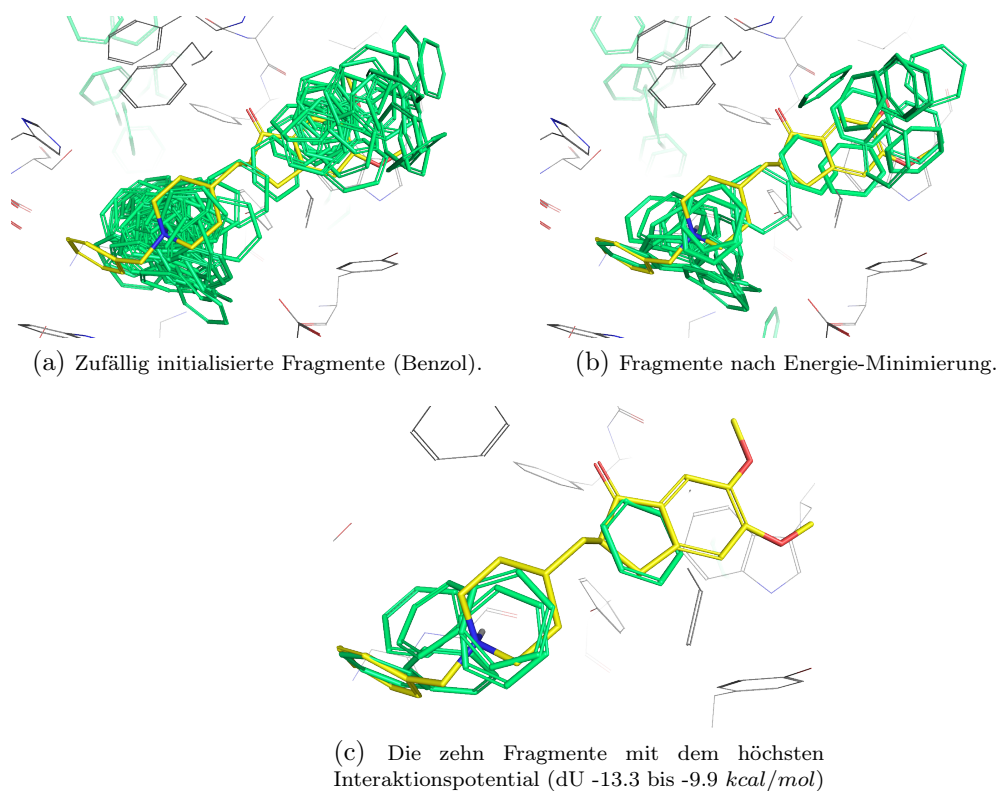


Abbildung 4.2: Methodik fragmentbasierter Ansätze am Beispiel von MOE-MFSS. Die Benzolfragmente (aromatisch-hydrophobe Eigenschaften) sind mit Abweichungen in der Lage, in Abwesenheit von Ligandinformation die aromatischen Funktionalitäten des Liganden (Donepezil) vorherzusagen (PDB 1eve; *Acetylcholinesterase*).

Kapitel 5

GRID-Hotspots im strukturbasierten Wirkstoffdesign

Hotspot-Information kann vielfältig in der Wirkstofffindung eingesetzt werden, z.B. im Rahmen von Datenbanksuchen (in einem Docking mit Nebenbedingungen, engl. *constrained docking* oder zur Ableitung strukturbasierter Pharmakophore)¹⁵¹. Alternativ kann diese Information für die Optimierung von Leitstrukturen verwendet werden. Nachfolgend wird exemplarisch an einigen Beispielen gezeigt, inwiefern die Berücksichtigung von GRID-*Hotspots* entscheidend zum Erfolg von strukturbasierten Ansätzen beigetragen hat. Ein Schwerpunkt ist hierbei auf Methoden gelegt worden, bei denen MIF-Informationen zur Erstellung von Pharmakophor-Modellen genutzt worden sind.

Es ist nachhaltig zu betonen, dass auch die anderen in Kapitel 4 besprochenen Methoden erfolgreich im Wirkstoffdesign verwendet worden sind. Dieses Kapitel beschränkt sich auf die GRID-Methode, da sie einen zentralen Teil CavKAs darstellt. Die hier vorgestellten Beispiele sind eher erfolgreiche Einzelanwendungen als eigenständige Methoden. Eine Auswahl an unterschiedlichen Methoden zur Pharmakophor-Ableitung wird in Kapitel 6 präsentiert. Eine Zusammenfassung über Methoden zur *Hotspot*-Detektion, die Verwendung von *Hotspot*-Information zur Erstellung strukturbasierter Pharmakophor-Modelle sowie weitere Anwendungsbeispiele geben BRENK und KLEBE¹⁵¹.

5.1 Erstellung strukturbasierter Pharmakophore

MASON und CHENEY nutzten MIF-Information für die Erstellung von Vierpunkt-Pharmakophoren und nutzten diese, um die Selektivität unterschiedlicher Serin-Proteasen vorherzusagen^{152,153}. Die Grundidee, dass alle essentiellen Merkmale in einem Vierpunkt-Pharmakophor vereint sind, wurde in der FLAP-Methode aufgegriffen und weiterentwickelt (s. Abschn. 7.6)¹⁵⁴.

FOX und HAAKSMA verwendeten GRID-MIFs, um in Abwesenheit von Ligandinformation strukturbasiert ein Thrombin-Pharmakophor zu erstellen¹⁵⁵. Die konturierten Felder wurden in ein Vierpunkt-Pharmakophor überführt. Die Autoren validierten ihre Methode erfolgreich in einem retrospektiven Screening mit Hilfe einer firmeneigenen Datenbank.

TINTORI et al. stellten mit ihrer *Targets Looking for Drugs*-Methode ein halb-automatisiertes Protokoll zur Ableitung strukturbasierter Pharmakophor-Modelle für Protein-Apo-Strukturen unter Verwendung von GRID-MIFs vor¹⁵⁶. Die Prozedur gliedert sich in drei Schritte: MIF-Berechnung, MIF-Konturierung und den Export der GRID-Minima in ein Catalyst-Pharmakophor-Model⁹².

Die Autoren empfahlen für die manuelle Platzierung und Priorisierung pharmakophorer Eigenschaften, die folgenden Kernaspekte zu berücksichtigen:

- die Interaktionsenergie (je negativer, desto besser),
- die Lokalisierung der Eigenschaften (je tiefer in der Bindetasche vergraben, desto besser),
- die Entfernung der Eigenschaften voneinander (min 1Å; max. 15Å voneinander entfernt; das letztgenannte Kriterium nur bei sehr großen Liganden),
- die Platzierung von H-Brücken-Donor- und -Akzeptor-Eigenschaften (in der Distanz einer H-Brückenbindung zu einer kompatiblen Aminosäureeigenschaft).

Die Leistungsfähigkeit der Methode wurde getestet, indem nach dem oben genannten Protokoll strukturbasierte Pharmakophore für die Zielstrukturen *Thioredoxin-Reduktase*, *HIV-1 Integrase* und *HIV-1 Reverse Transkriptase* erstellt wurden.

Für *HIV-1 Reverse Transkriptase* wurde zusätzlich die Rezeptorflexibilität durch die Berechnung von Molekül-Dynamiken (Abk. MD) simuliert und Pharmakophor-Modelle für unterschiedliche MD-Konformere abgeleitet. Die erstellten Modelle

wurden zum Vorfiltern der ASINEX-GOLD-Datenbank¹⁵⁷ verwendet und die den Pharmakophor erfüllenden Moleküle anschließend in einem Docking weiter analysiert. Eine Auswahl an Molekülen wurde schließlich in einem biochemischen Assay getestet. Für alle drei untersuchten Zielstrukturen konnten so prospektiv neue Inhibitoren gefunden werden.

5.2 Leitstruktur Optimierung

Eine weitere Möglichkeit *Hotspot*-Information zu nutzen, besteht in der Modifikation von bereits bekannten aktiven Substanzen. Durch kleinere Veränderungen am molekularen Grundgerüst wird versucht, den Liganden besser an seine Zielstruktur anzupassen, um potentere Substanzen zu erhalten. Dieses Vorgehen wird auch als Leitstruktur-Optimierung (engl. *lead optimization*) bezeichnet.

Die Entwicklung des *Neuraminidase*-Inhibitors Zanamivir (Relenza[®]) stellt die wohl bekannteste prospektive Anwendung GRIDs im strukturbasierten Wirkstoffdesign dar¹⁵⁸. GRID-*Hotspots* inspirierten dazu, eine Hydroxyl-Gruppe eines relativ schwachen Inhibitors gegen eine basische funktionelle Gruppe zu ersetzen. Ein daraufhin synthetisiertes Guanidino-substituiertes Derivat (Zanamivir) war in der Lage, die *Neuraminidase* im niedrig nanomolaren Bereich zu inhibieren.

Kapitel 6

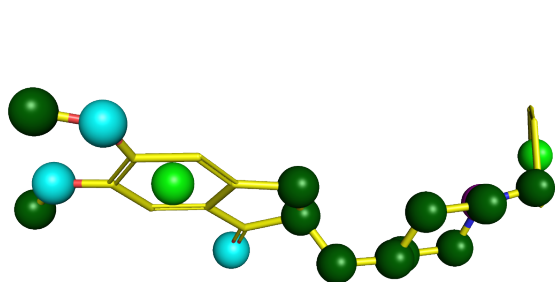
Methoden zur strukturbasierten Pharmakophor-Ableitung

Nachfolgend wird eine Auswahl von Programmen zur Ableitung strukturbasierter Pharmakophor-Modelle vorgestellt. Die Methoden unterscheiden sich zum einen bezüglich der Informationsbasis, die der Ableitung des Modells zugrunde liegt (Ligand-Rezeptor-Komplex oder Protein-Apo-Struktur) und zum anderen durch die Kriterien, die verwendet werden, um wichtige Schlüsseleigenschaften zu identifizieren (geometrische, energetische Kriterien und kombinierte Ansätze). Alle Methoden verbindet, dass vorgeschlagene pharmakophore Eigenschaften als *Modellvorschläge* zu verstehen sind und gegebenenfalls weiterer Modifikationen bedürfen.

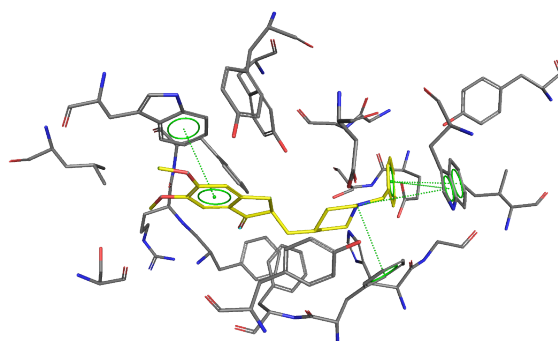
Bei Methoden, die auf Basis einer Apo-Struktur erstellt werden, ist ein größerer Einfluss des Anwenders erforderlich als bei Methoden, die Ligand-Rezeptor-Komplexe als Informationsbasis verwenden (sogenannte 'komplexbasierte Methoden'). Methoden zur *Hotspot*-Detektion definieren die Lage potentieller Eigenschaften vage und bieten somit relativ viel Spielraum für die Platzierung pharmakophorer Sphären, was klar als Nachteil zu werten ist.

6.1 Komplexbasierte Methoden: Geometrische Kriterien

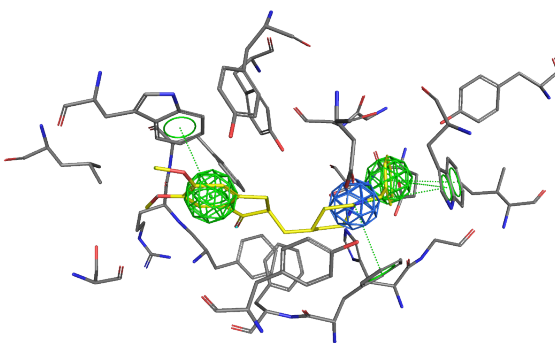
In den meisten Fällen wird auf der Basis von Ligand-Rezeptor-Komplexen nach geometrischen Kriterien ermittelt, ob pharmakophore Ligandeigenschaften mit der Bindetasche interagieren, um diese in einem Pharmakophor-Modell zu berücksichtigen. Abbildung 6.1 zeigt die schematische Vorgehensweise.



(a) Definition pharmakophorer Ligandeigenschaften. Hellgrün: aromatisch; dunkelgrün: hydrophob; türkis: H-Brücken-Akzeptor.



(b) Detektion von Ligand-Rezeptor-Interaktionen mittels geometrischer Kriterien (hellgrün gepunktete Linien).



(c) Interagierende Ligandeigenschaften werden in ein Pharmakophor-Modell übersetzt.



(d) Interaktionsbasiertes Pharmakophor-Modell, das zur Datenbanksuche verwendet werden kann.

Abbildung 6.1: Ableitung strukturbasierter Pharmakophore aus Ligand-Rezeptor-Komplexen auf Basis geometrischer Interaktionskriterien (PDB: 1eve; *Acetylcholinesterase*; Ligand: Donepezil).

Zunächst werden Ligandfunktionalitäten in pharmakophore Eigenschaften übersetzt (s. Abb. 6.1a). Im sich anschließenden Schritt wird ermittelt, welche der Ligandeigenschaften an einer Interaktion mit der Bindetasche beteiligt sind (s. Abb. 6.1b und Abb. 6.1c). Hierzu werden zunächst sinnvolle Eigenschafts-Kombinationen zwischen Ligand und Bindetasche gesucht sowie Distanzen und Winkel abgeglichen, um im Anschluss nach geometrischen Regeln zu prüfen, ob es zu einer Interaktion kommen kann. In Kapitel 11 erfolgt die detaillierte Beschreibung, nach welchen geometrischen Gesetzmäßigkeiten die Ligand-Rezeptor-Interaktionsanalyse in CavKA implementiert worden ist. Interagierende Ligandeigenschaften werden nun in ein Pharmakophor-Modell übersetzt, das für Datenbanksuchen verwendet werden kann (s. Abb. 6.1d).

6.1.1 Ligandscout

Ligandscout ist derzeit das wohl bekannteste und am weitesten verbreitete Programm zur komplexbasierten Pharmakophor-Ableitung, das nach den oben genannten Prinzipien arbeitet⁸⁸. Die geometrischen Kriterien zur Interaktionserkennung, wie Distanzen und Winkel, können vom Anwender editiert und angepasst werden. Eine Besonderheit ist, dass Ligandscout die Bindungsordnung im cokrystallisierten Liganden überprüft und falls notwendig korrigiert¹⁵⁹. Dies ist insofern relevant, da in den verwendeten Strukturen der Protein-Datenbank (Abk. PDB)^{160,161} im .pdb Format meist keine Bindungsordnung für kleine Moleküle definiert ist. Aus falsch interpretierten Ligandeigenschaften können sehr schnell fehlerhafte Pharmakophor-Modelle resultieren, weshalb gebundene Liganden stets zusätzlich visuell geprüft werden sollten. Zusätzlich bietet die aktuelle Ligandscout-Software die Möglichkeit, ligandbasiert Pharmakophor-Modelle zu erstellen und besitzt eine eigene Plattform für Datenbanksuchen¹⁶².

Neben der oben genannten Methode bieten mittlerweile viele Modelling-Programme, wie z.B. Accelrys Discovery Studio⁹² und Phase⁶³, die Möglichkeit komplexbasierte Pharmakophore nach geometrischen Regeln abzuleiten.

6.2 Komplexbasierte Methoden: Energetische Kriterien

6.2.1 E-Pharmacophores (Phase)

SALAM et al. entwickelten mit ihrem E-Pharmakophor Ansatz eine Methode, die Ligandeigenschaften in der Bindetasche über die Scoring-Funktion eines Docking-Programms (*Glide XP*¹⁶³) bewertet. Hierbei werden jeder pharmacophoren Ligandeigenschaft der an dieser entsprechenden Position ermittelte *Glide-XP-Score* zugeordnet. Anhand der Scores werden die Eigenschaften gewichtet¹⁶⁴. Die Autoren erarbeiteten auf dieser Grundlage einen empirischen Regelsatz, der automatisierte Pharmacophor-Ableitungen ermöglicht:

- Eine Eigenschaft wird als essentiell (engl. *explicit matching*) definiert, wenn an der entsprechenden Eigenschafts-Koordinate ein Score < -1.0 kcal/mol registriert worden ist.
- Bei einem Pharmakophor-Modell, das aus drei bis vier Sphären besteht, muss das Modell von einem Liganden mindestens in drei Fällen räumlich und der Eigenschaft nach erfüllt sein (engl. *partial matches*).
- Bei einem Pharmakophor-Modell, das aus fünf oder mehr Sphären besteht, muss das Modell von einem Liganden mindestens in vier Fällen räumlich und der Eigenschaft nach erfüllt sein.
- In einem Pharmakophor-Modell werden maximal sechs Sphären definiert.

Mit dem E-Pharmakophor-Ansatz erzielten die Autoren in ihrer retrospektiven Validierstudie höhere Anreicherungen aktiver Moleküle als mit einem nach geometrischen Regeln abgeleiteten Pharmakophor-Modell (engl. *contact-based-pharmacophor*; Abk. C-Pharmakophor).

6.2.2 GBPM (GRID Based Pharmacophore Model)

Das von ORTUSO et al. entwickelte GRID Based Pharmacophore Model (Abk. GBPM) verwendet GRID-MIFs, um automatisiert und ohne Einfluss des Anwenders Pharmacophor-Modelle aus einem Ligand-Rezeptor-Komplex oder einem Protein-Protein-Interface abzuleiten^{165,166}. Bei der Methode werden zunächst MIFs des Ligand-Rezeptor-Komplexes, des Proteins sowie des Liganden berechnet. Aus den drei Feldern werden die gemeinsamen Areale identifiziert und konturiert.

Die Autoren verwendeten die Dry-, O- und N1-Sonde und empfahlen zur Konturierung einen Schwellenwert von 5-15% über dem jeweils globalen Minimum der verwendeten Sonde. Erhaltene *Hotspots* wurden in ein Catalyst-Pharmakophor⁹² exportiert und anhand der Interaktionsenergien gewichtet. Die Methode wurde von den Autoren erfolgreich in einem retrospektiven Virtuellen Screening an zwei Protein-Protein-Interfaces validiert (*X-linked inhibitor of apoptosis* und *Interleukin-8-Dimer*).

6.3 Komplexbasierte Methoden: Geometrische und energetische Kriterien

6.3.1 Pocket 2.0

Die von CHEN und LAI entwickelte Methode Pocket 2.0¹⁶⁷ interpretiert Ligand-Rezeptor-Komplexe geometrisch und nutzt den gitterbasierten SCORE-Ansatz¹⁴⁴, um interagierende Ligandeneigenschaften zu gewichten und diese adäquat in einer Pharmakophor-Hypothese zu berücksichtigen. Die Autoren verwendeten eine H-Brücken-Donor-, eine H-Brücken-Akzeptor- und eine hydrophobe Sonde für die Berechnung der Interaktionspotentiale. Die Konturierung der SCORE-Felder erfolgte empirisch. Die Methodvalidierung bestand in der Reproduktion von bereits publizierten Pharmakophor-Modellen der Zielstrukturen *Cyclinabhängige-Kinase 2*, *HIV-1 Protease*, *17 β -Hydroxysteroid-Dehydrogenase* und dem *Estrogen-Rezeptor*. Optional lässt sich Pocket 2.0 für das Erstellen strukturbasierter Pharmakophore auf Basis von Apo-Strukturen verwenden.

Von allen in Kapitel 6 vorgestellten Methoden weisen CavKA und Pocket 2.0 konzeptionell die größten Gemeinsamkeiten auf. Obwohl Pocket 2.0 laut Publikation auf Anfrage kostenlos über die Autoren erhältlich sein soll, blieben schriftliche Anfragen unbeantwortet. Ein direkter Vergleich von CavKA mit Pocket 2.0 im Rahmen eines Virtuellen Screenings war somit nicht möglich.

6.4 Methoden für Apo-Strukturen

Das computergestützte Ableiten von strukturbasierten Pharmakophor-Modellen auf der Basis von Apo-Strukturen in automatisierter Form stellt eher eine Ausnahme dar. Entsprechend wenig ist auf diesem Gebiet publiziert und verfügbar (vgl. Pocket 2.0 und *Targets Looking for Drugs*, vgl. Abschn. 5.1).

Ist nur eine Apo-Proteinstruktur verfügbar, muss zunächst ein Bindetaschenareal definiert werden, in dem ein mutmaßlicher Ligand bindet, um dort pharmakophore Sphären zu definieren. Für diese Bindetaschendefinition kann entweder auf Literaturwissen zurückgegriffen (falls vorhanden) oder ein Programm zur Bindetaschendetektion verwendet werden. Diese Programme verwenden meist geometrische oder energetische Kriterien, um eine potentielle Bindetasche zu definieren. HENRICH et al. geben eine Zusammenfassung über gebräuchliche Methoden zur Bindetaschendetektion in einem Übersichtsartikel¹⁶⁸.

Ist eine Bindetasche definiert worden, kann mit kraftfeld-, bzw. wissensbasierten Methoden nach *Hotspots* gesucht werden, die für die Definition pharmakophorer Sphären verwendet werden können. Die Erstellung der Pharmakophor-Modelle erfolgt meist manuell und wird an die jeweilige Zielstruktur angepasst. Hierbei ist der Übergang von einer individuellen Fallstudie (s. *Targets Looking for Drugs*) über halbautomatische Methoden bis hin zum automatisierten Programm für *de-novo* Design fließend.

Im folgenden Abschnitt wird die Methode Mr. Brightsite als halbautomatisches Protokoll zur Pharmakophor-Ableitung auf Basis von Apo-Strukturen vorgestellt. Hierbei werden wissensbasierte Methoden der *Hotspot*-Erkennung mit gitterbasierten kombiniert. Im Anschluss wird anhand der LUDI-Methode exemplarisch erläutert, wie die Software für das *de-novo* Design von Substanzen genutzt werden kann.

6.4.1 Mr. Brightsite

Inspiziert von FOX und HAAKSMA¹⁵⁵ sowie *Targets Looking for Drugs*¹⁵⁶ (s. Abschn. 5.1) wurde das Programm Mr. Brightsite als Vorläufer von CavKA entwickelt. Ziel war es hierbei die Kerngedanken von *Targets Looking for Drugs* aufzugreifen und eine Methode zur Ableitung von Pharmakophor-Modellen auf Basis von Apo-Strukturen zu implementieren.

Mr. Brightsite (s. Abb. 6.2) kombiniert dabei wissensbasierte und empirische Ansätze, um *Hotspots* in Proteinbindetaschen in Abwesenheit von Ligandinformation zu detektieren und auf Basis dieser Information Pharmakophor-Modelle zu erstellen. Die Bindetaschendetektion wurde über die Webinterfaces von Pocketfinder¹⁶⁹ (Ligsite-Algorithmus¹⁷⁰) und Q-Site Finder¹⁷¹ (Qsite-Algorithmus¹⁷²) durchgeführt. Als Bindetaschendefinition (und somit die Definition des Areal für die nachfolgenden *Hotspot*-Berechnungen) wurden die vereinigten Koordinaten der größten von der jeweiligen Methode gefundenen Kavitäten verwendet. Konturierte GRID-MIFs wurden als empirisch-energetischer Ansatz zur *Hotspot*-Detektion verwendet. Analog zu den geometrischen Regeln von Cavbase¹¹⁶ wurden *cavity flanking residues* identifiziert, die zur Definition pharmakophorer Proteineigenschaften genutzt wurden. Als *Hotspots* für die manuelle Pharmakophor-Erstellung wurden jene GRID-MIFs betrachtet, die sich in der Nähe einer komplementären Proteineigenschaftsoberfläche befanden.

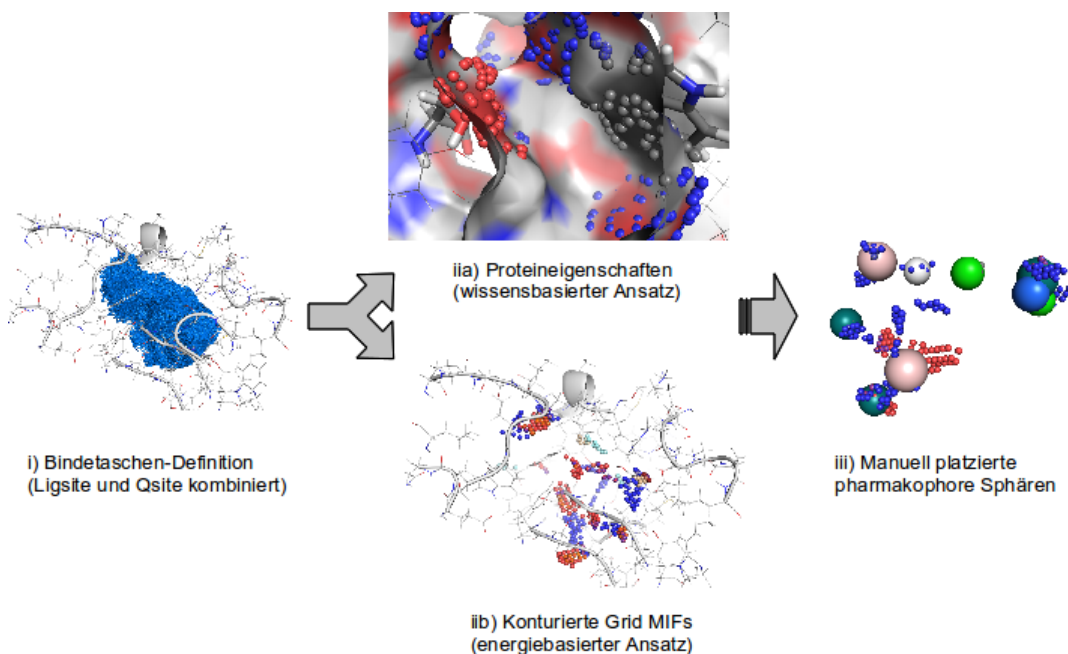


Abbildung 6.2: Schematische Darstellung von Mr. Brightsite am Beispiel der *Acetylcholinesterase* (PDB 1eve). Erklärung siehe Text.

Obwohl im Einzelfall gute Ergebnisse mit der Methode erzielt worden sind, ist der Ansatz nicht weiter verfolgt worden. Es ließen sich keine objektiven Regeln ermitteln, in welcher Distanz sich Proteineigenschaft und konturiertes MIF zueinander befinden müssen, um einen *Hotspot* zu definieren. Abbildung 6.3 verdeutlicht, dass MIFs auch bei starken Konturierungsgraden (z.B. 20% der Punkte mit dem höchsten Interaktionspotential) oft beträchtliche Ausmaße annehmen können. Man kann in diesen Fällen eher von *Hotplanes* (also 'heißen Flächen') als von *Hotspots* sprechen. Ein weiteres Problem der MIFs besteht darin, dass sie enorm rauschanfällig sind: Das MIF der Bindetasche zeigt die maximalen Interaktionswerte (grüne Sphären) an einer anderen Stelle als die Interaktion im Ligand-Rezeptor-Komplex tatsächlich zu beobachten ist (gestrichelte Linie). In Übereinstimmung mit *Targets Looking for Drugs* lässt sich zusammenfassend schließen, dass GRID-Felder weniger eine strikte räumliche Begrenzung für pharmakophore Eigenschaften, als eher richtungsweisende Indikatoren für deren Platzierung darstellen¹⁵⁶.

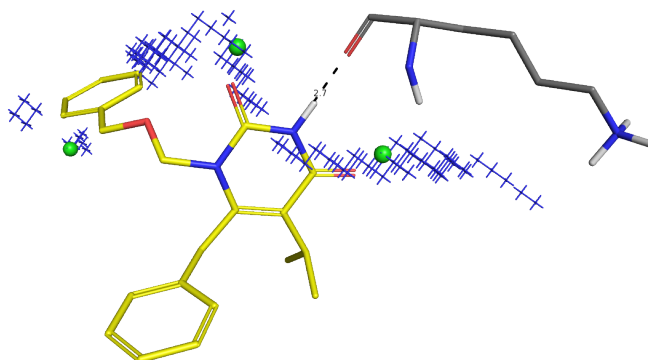


Abbildung 6.3: Ligand-Rezeptor-Komplex von *HIV-Reverse-Transkriptase* (PDB 1jla) mit einem konturierten N1-MIF (20% der Punkte mit dem höchsten Interaktionspotential; blaue Kreuze). Die N1-Punkte mit der höchsten Interaktionsenergie im jeweiligen Feld-Abschnitt sind als grüne Sphäre dargestellt. Der entscheidende H-Brücken-Donor (gestrichelte Linie) wird nur am Rand des N1-MIFs touchiert.

Die manuelle Platzierung pharmakophorer Sphären auf diesen Flächen ist alles andere als eindeutig. Zum hohen manuellen Aufwand des Anwenders für die Modell-Erstellung kommt die nicht eindeutige *Hotspot*-Erkennung hinzu. Interaktionsenergien allein liefern nur begrenzt verwertbare Informationen. Aufgrund der individuellen Interpretation der *Hotspots* ist die Methode enorm anfällig für eine 'Verzerrung' durch den Anwender. Mr. Brightsite ist somit vielmehr als ein Werkzeug für den medizinischen Chemiker zu verstehen als ein vollautomatisiertes Programm zur Pharmakophor-Ableitung.

6.4.2 LUDI (de-novo Design)

Die oben beschriebenen Programme zur Ableitung strukturbasierter Pharmakophore in Abwesenheit von Ligandinformation sind klar von *de-novo* Design-Programmen, wie z.B. LUDI^{173,174} (derzeit implementiert in Discovery Studio⁹²), abzugrenzen. Anstatt Modelle für eine Datenbanksuche abzuleiten, steht beim *de-novo* Design die grundlegende Neuentwicklung von Substanzen im Vordergrund.

BOEHM gibt in einer Übersichtsarbeit eine Zusammenfassung über die Entwicklung des Computerprogramms, der Scoring-Funktion sowie erfolgreiche Anwendungen¹⁷⁴. LUDI verwendet für das *de-novo* Design ein dreischrittiges System: Zunächst werden Positionen in der Bindetasche ermittelt, an denen eine potentielle Ligand-Rezeptor-Interaktion stattfinden kann. Diese sogenannten *interaction sites* können wissensbasiert (über statistische Häufigkeitsverteilungen von Interaktionen in der CSD¹⁷⁵; vgl. Abschn. 4.1), regelbasiert¹⁷⁶ ^I oder alternativ über GRID-MIFs¹²⁹ ^{II} ermittelt werden. Im zweiten Schritt der Methode werden Molekülfragment-Anker in einer Datenbank gesucht, die zur Bindetasche kompatible Interaktionen formen können und auf den *interaction sites* platziert werden. Im letzten, sich anschließenden Schritt werden die Anker-Fragmente über Brücken-Fragmente miteinander verbunden, um so neue Liganden zu entwickeln, die über eine Scoring-Funktion bewertet werden.

^IBoehms Regelsatz selbst ist auf Basis statistischer Analysen von nicht-kovalenten Interaktionen in der CSD entwickelt worden¹⁷⁶.

^{II}In der ursprünglichen LUDI-Version¹⁷³ konnten GRID-MIFs¹²⁹ anstatt des regelbasierten Ansatzes verwendet werden, um *interaction sites* für die Platzierung der Molekülfragment-Anker zu definieren.

Kapitel 7

Ligandzentrische Ansätze

Ligandzentrische Ansätze nehmen eine Zwischenstellung zwischen den ligand- und strukturbasierten Methoden im Virtuellen Screening ein. Idealerweise wird der co-kristallisierte Ligand eines Ligand-Rezeptor-Komplexes als Referenz für Datenbank-suchen verwendet, da in vielen Fällen davon ausgegangen wird, dass dieser den nativen Bindungsmodus am besten widerspiegelt und zusätzlich Proteininformation enthält, da das Protein die Gestalt und die Orientierung des Liganden in der Bindetasche mit beeinflusst. Alternativ dazu kann auch ein energetisch günstiges Konformer oder eine 3D-Struktur verwendet werden.

Bei ligandzentrischen Methoden wird Ligandinformation über unterschiedliche dreidimensionale Deskriptoren beschrieben. Eine mögliche Eigenschaft, die zur Repräsentation des Liganden verwendet werden kann, ist das Ligandvolumen (engl. *shape*), weshalb solche Ansätze auch als *shape-basiert* bezeichnet werden. Die meisten *shape*-basierten Ansätze haben gemeinsam, dass die schweren Ligand-Atome (also alle Atome außer Wasserstoff) durch das Integral dreidimensionaler Gauß-Peaks (also durch Gauß-Volumina) beschrieben werden statt durch klar definierte pharmakophore Sphären (sogenannte Harte-Sphären)¹⁷⁷. Beim Vergleich von zwei Molekülen wird die Überlappung der Gauß-Volumina maximiert und bewertet¹⁷⁸. Die Repräsentation eines Liganden durch Gauß-Volumina besitzt den großen Vorteil, dass Information sehr viel 'weicher' als bei harten Sphären codiert wird: Im Gegensatz zu den Harten-Sphären-Modellen wird nicht binär klassifiziert, ob der Pharmakophor erfüllt ist oder nicht (vgl. Abschn. 2.1). Vielmehr wird das Maß an Überlappung zwischen Referenz- und Datenbank-Molekül über einen Score bewertet, was zu einer Hitliste mit kontinuierlichen geordneten Scores führt.

Shape-basierte Ansätze sind bereits in vielen Programm-Paketen (*Cat-Shape*, Discovery Studio⁹²; *Phase-Shape* in Phase⁶³) implementiert sowie als einzelne Programme erhältlich (*ROCS_{shape}*, PARASHIFT¹⁷⁹, HEX¹⁸⁰, USR¹⁸¹, PAPER¹⁸², Piramid¹⁸³).

Ligand-Rezeptor-Interaktion wird jedoch keineswegs nur über die Molekülgestalt, respektive das molekulare Volumen, vermittelt. Die pharmokophoren Eigenschaften spielen eine mindestens ebenso wichtige Rolle. Daher werden Deskriptoren pharmakophorer Eigenschaften häufig mit *shape*-Deskriptoren kombiniert. Um pharmakophore Eigenschaften zu modellieren werden unterschiedliche Ansätze genutzt, die feldbasierte Methoden (Fieldscreen¹⁸⁴, FLAP⁴⁵), Elektrostatik (shaEP¹⁸⁵, EON¹⁸⁶) und auf Substruktursuche-basierende Methoden (Pharao¹⁸⁷, *ROCS_{color}*) umfassen^I.

Besonders hervorzuheben ist, dass bei ligandzentrischen Ansätzen keine aufwändige Protein-Vorbereitung (wie z.B. beim Docking) erforderlich ist. So muss beispielsweise bei der Verwendung des OpenEye Docking Programms FRED¹⁸⁹ ein inneres Bindetaschen Volumen (in dem der gedockte Ligand liegen *muss*) und ein äußeres Volumen (in dem der gedockte Ligand liegen *kann*) definiert werden. Die Ausrichtung von Wasserstoffatomen sowie Protonierungszustände müssen geprüft und ggf. angepasst werden, damit entstehende Interaktionen von der Scoring-Funktion adäquat bewertet werden können. Man kann sich leicht vorstellen, dass alle diese Parameter einen immensen Einfluss auf das Docking-Ergebnis haben können und eine erhebliche Intervention des Anwenders erfordern.

Bei allen ligandzentrischen Methoden, die pharmakophore Eigenschaften modellieren, ist hingegen zu beachten, dass Referenz- und Datenbank-Molekül Protonierungszustände aufweisen müssen, die dem gleichen pH-Wert entsprechen, um eine Vergleichbarkeit zu ermöglichen. Ein Angleich auf den physiologischen pH-Wert von 7.4 wird bei ROCS und EON automatisch vorgenommen. Nachfolgend werden ligandzentrische Ansätze vorgestellt, die im Rahmen dieser Arbeit verwendet worden sind.

^IÜber Substruktursuchen¹⁸⁷ werden regelbasiert¹⁸⁸ pharmakophore Ligandeigenschaften identifiziert.

7.1 ROCS

ROCS¹⁹⁰ ist das Akronym für *Rapid Overlay of Chemical Structures* und wird von vielen Anwendern als der Industriestandard für ligandzentrische Ansätze betrachtet¹⁹¹. Im direkten Vergleich mit Docking-Methoden schnitt ROCS vergleichbar oder besser ab^{192,193}, bewies die Fähigkeit biologisch aktive Moleküle aus einem Validier-Datensatz anzureichern¹⁹⁴ und wurde darüber hinaus erfolgreich in prospektiven Studien verwendet^{195–197}.

Der Anwender hat die Möglichkeit unterschiedliche Scoring-Funktionen für den Vergleich von Ligand und Datenbank-Molekül heranzuziehen: das Ligandvolumen allein ($ROCS_{shape}$), pharmakophore Eigenschaften ($ROCS_{color}$) oder eine Kombination aus beiden ($ROCS_{combo}$). Der *Combo Score* gewichtet die Eigenschaften *shape* und *color* im gleichen Verhältnis und wird in den Standardeinstellungen verwendet. KIRCHMAIR et al. untersuchten unterschiedliche Gewichtungen und empfahlen ein Verhältnis von $1(shape): 4(color)$ ¹⁹¹.

Analysen von KIRCHMAIR et al. zeigten, dass statt eines co-kristallisierten Liganden auch ein energetisch günstiges Konformer (engl. *low energy conformer*) als Referenz verwendet werden kann, ohne dass in einem Screening ein nennenswerter Verlust an Leistungsfähigkeit zu beobachten ist¹⁹¹. In einer anderen Studie erzielten KIRCHMAIR et al. auch bei einer sehr niedrigen Auflösung des verwendeten Konformer-Ensembles annähernd gleich gute Ergebnisse wie mit einer deutlich höheren Auflösung¹⁹⁸. Diese Beobachtungen führten die Autoren auf die Fähigkeit des verwendeten Konformerengenerators zurück (in diesem Falle OMEGA¹⁹⁹) Konformere zu berechnen, die dem nativen co-kristallisierten Liganden sehr ähnlich sind¹⁹⁸.

7.2 EON

EON²⁰⁰ verwendet elektrostatische Potentiale für die Definition pharmakophorer Eigenschaften, da räumliche Gestalt und Elektrostatik als die Haupteigenschaften für molekulare Erkennung gelten¹⁸⁶. EON selbst überlagert keine Moleküle. Es verwendet bereits durch ROCS superpositionierte Moleküle und berechnet anschließend elektrostatische Potentiale. Auch hier hat der Anwender die Auswahl an unterschiedlichen Scoring-Funktionen, die *shape* (identisch mit $ROCS_{shape}$), elektrostatische Potentiale oder eine Kombination aus beiden Eigenschaften umfassen.

7.3 Piramid

Piramid ist eine frei verfügbare Software, die analog zu *ROCS_{shape}* die Gestalt von Referenz- und Datenbank-Molekül abgleicht, wobei schwere Atome ebenfalls durch dreidimensionale Gauß-Peaks repräsentiert werden und die Überlappung zwischen den zu vergleichenden Molekülen maximiert wird¹⁸³. In der aktuell verfügbaren Version besteht keine Möglichkeit pharmakophore Eigenschaften in der Bewertung zu berücksichtigen.

Aus rechtlichen Gründen wurde Piramid im April 2012 in 'shape-it' umbenannt²⁰¹. Am Programm selbst sind keine Änderungen vorgenommen worden²⁰².

7.4 Pharao

Pharao¹⁸⁷ ist das Akronym für *Pharmacophore Alignment and Optimization*. Das frei verfügbare Programm besitzt zahlreiche Gemeinsamkeiten mit *ROCS_{color}*. Ligand-Funktionalitäten werden in pharmakophore Ligandeigenschaften übersetzt, die im Gegensatz zu konventionellen Pharmakophor-Suchen durch einen dreidimensionalen Gauß-Peak repräsentiert werden. Im anschließenden Optimierungsschritt wird die Überlappung zwischen Pharmakophor-Modell und Datenbank-Molekül maximiert, sodass möglichst viele Eigenschaften *räumlich* und der *pharmakophoren Eigenschaften nach* zur Deckung gebracht werden^{187,203}. Weitere Details zu Pharao, insbesondere zur Eigenschafts-Definition, finden sich in Kapitel 10.

Aus rechtlichen Gründen wurde Pharao im April 2012 in 'align-it' umbenannt²⁰⁴. Am Programm selbst sind keine Änderungen vorgenommen worden²⁰².

7.5 ShaEP

ShaEP¹⁸⁵ ist die Abkürzung für *Shape and Electrostatic Potential*. Die Methode abstrahiert die zu untersuchenden Liganden in einem mathematischen Graphen (engl. *field graph*). Nach geometrischen Regeln (abhängig vom Hybridisierungsgrad der Ligandatome) werden zunächst Oberflächenpunkte berechnet. Jeder dieser Punkte wird mit einem Deskriptorwert für die lokale Umgebung versehen. Dieser umfasst ein Histogramm der Distanzen vom Punkt des Ligandatoms zu der van-der-Waals Oberfläche des Liganden sowie einen Wert für das elektrostatische Potential an dieser Stelle.

Die so erhaltenen Punkte mit den zugehörigen Werten sowie die euklidische Distanz zwischen den Punkten stellen den Graphen dar und repräsentieren das ursprüngliche Molekül. Mittels *Clique Detection* wird der maximal gemeinsame Subgraph für Referenz- und Datenbank-Molekül ermittelt. Anschließend wird die Überlappung von molekularem Volumen und Elektrostatik maximiert und die finale Superpositionierung bewertet. Als Bewertungsfunktionen stehen $ShaEP_{shape}$ (nur die Überlappung der Molekülvolumina berücksichtigend) und $ShaEP_{combo}$ (Elektrostatik und molekulares Volumen betrachtend) zur Verfügung. Im Rahmen des Optimierungsprozesses werden analog zu ROCS schwere Ligandatome als dreidimensionale Gauß-Peaks repräsentiert. Oberflächenpunkte werden ebenfalls als Gauß-Peaks betrachtet und mit dem elektrostatischen Potential gewichtet. ShaEP ist kostenlos auf Anfrage erhältlich.

7.6 FLAP

FLAP ist das Akronym für *Fingerprints for Ligands and Proteins*^{45,154}. Obwohl die Methode nicht im Rahmen der ligandenzentrischen Analysen in dieser Arbeit verwendet worden ist (die Software ist derzeit nicht kommerziell erhältlich und wird nur ausgewählten Arbeitsgruppen zur Verfügung gestellt; Stand: 15.11.2011), wird sie der Vollständigkeit halber aufgeführt, da GRID-MIFs¹²⁹ zur Definition pharmakophorer Eigenschaften verwendet werden. FLAP nimmt eine Sonderstellung ein, da es sowohl als ligandenzentrische als auch als strukturbasierte Screening Methode verwendet werden kann.

Das Grundprinzip der Methode ist die Repräsentation kleiner Moleküle durch Vierpunkt-Pharmakophore, die in einem Fingerprint codiert werden. Das Konzept solcher Pharmakophor-Fingerprints ist dabei nicht neu²⁰⁵ II: MASON und CHENEY verwendeten Pharmakophor-Quadrupletts unter Verwendung von GRID-MIFs um Serin-Proteasen zu klassifizieren^{207,208}. Quadrupletts codieren im Gegensatz zu Tripletts auch Chiralität. Vorteile solcher Fingerprints sind, dass sie extrem speichereffizient sind und sehr schnell miteinander abgeglichen werden können⁷⁶. Sie eignen sich somit auch für Screenings auf sehr großen Datensätzen⁷⁶.

^{II}Ursprünglich verwendeten MASON et al. SMARTS²⁰⁶-Substruktursuchen für die Definition pharmakophorer Eigenschaften.

Es werden zunächst von einem Liganden oder einer Bindetasche MIFs unterschiedlicher Sonden berechnet (für jeweils unterschiedliche pharmakophore Eigenschaften), von denen anschließend nur lokale Minima analysiert werden. Ein solches Minimum (sog. MINI-Punkt) wird definiert als Gitterpunkt, der ausschließlich von Gitterpunkten mit einem positiveren Interaktionspotential umgeben ist²⁰⁹. Jeweils vier solcher MINI-Punkte repräsentieren ein Vierpunkt-Pharmakophor.

Die Besonderheit der FLAP Methode besteht darin, dass die Information des gesamten Vierpunkt-Pharmakophors in einem elfdimensionalen Bitstring codiert wird (vier Bits für die Eigenschafts Typen, sechs Bits für die Inter-Eigenschafts-Distanzen in diskreter Form^{III} und einem Bit, das der Summe der Interaktionsenergien aller Eigenschaften entspricht⁴⁵). Da für jede Zielstruktur mitunter sehr viele Minima existieren können, empfehlen die Autoren diese auf 25-30 Minima pro Sonde⁴⁵ (mit dem höchsten Interaktionspotential) zu begrenzen. Die Methode untersucht nun alle möglichen Vierpunkt-Kombinationen von Referenz- und Datenbank-Molekül. Vierpunkt-Pharmakophore können für Liganden und Proteine berechnet werden, woraus sich eine Vielzahl von denkbaren Anwendungsgebieten ergibt (jeweils in Klammern):

- Ligand - Ligand Ähnlichkeitsanalyse (ligandzentrischer Screeningansatz; QSAR-Analysen)
- Ligand - Protein Ähnlichkeitsanalyse ('automatisiertes' Docking; QSAR-Analysen)
- Protein - Protein Ähnlichkeitsanalyse (Bindetaschenvergleiche; Bindetaschenklassifikation; QSAR-Analysen)

SCIABOLA et al. veröffentlichten erfolgreiche Anwendungsbeispiele²⁰⁹. Des Weiteren verglichen CROSS et al. die FLAP Methode u.a. gegen die Fieldscreen-Methode¹⁸⁴ und erzielten beeindruckende Ergebnisse¹³³. Hierbei wurde der auch in dieser Arbeit verwendete Fieldscreen-Datensatz zur Validierung der Methode genutzt.

^{III}Es werden nicht die absoluten Euklidischen Distanzen berücksichtigt, sondern Distanzbereiche (z. B. zwischen 1-2Å, 3-4Å usw.).

Kapitel 8

Flexibilität von Datenbank-Molekülen

Wenn mit einem 3D-Modell eine Molekül-Datenbank durchsucht wird, müssen die Moleküle ebenfalls in dreidimensionaler Form vorliegen. Auch wenn es theoretisch denkbar ist, jedes Datenbank-Molekül durch ein einzelnes, energetisch günstiges Konformer zu repräsentieren, werden bei Verwendung einer multikonformeren Datenbank häufig deutlich bessere Screening-Ergebnisse erzielt²¹⁰.

Konformere werden genutzt, um Ligandflexibilität zu simulieren. Sie können entweder vorher durch ein externes Programm berechnet werden (engl. *rigid-body-treatment*), während der Datenbanksuche erzeugt und optimiert werden (engl. *on-the-fly*) oder ein bereits energetisch günstiges Konformer wird optimiert (engl. *directed-tweak*)⁷⁶.

Vorteile des *rigid-body-treatments* sind die Wiederverwertbarkeit der Konformere (die nur einmalig generiert werden müssen), jedoch auf Kosten eines hohen Speicherplatzbedarfs. Werden Konformere *on-the-fly* generiert, muss dies bei jedem Screening getan werden, was die Rechenzeit enorm erhöht, aber auch deutlich weniger Speicherkapazität benötigt. Im Gegensatz zum *rigid-body-treatment* können bei jeder Rechnung unterschiedliche Lösungen erhalten werden.

WOLBER et al. merkten an, dass bei der *directed-tweak* Methode der Konformenraum bereits stark reduziert ist. Bei zusätzlicher flexibler Optimierung besteht daher die Gefahr, dass statt des globalen Konformenraums eher lokale Minima abgebildet werden¹⁶².

Je nachdem, wie Konformere generiert worden sind, können Screening-Ergebnisse voneinander abweichen²¹¹. Der Erstellung von Konformeren-Ensembles sollte somit stets ein besonderes Augenmerk gelten. BOSTROEM²¹² und SCHWAB²¹³ thematisieren in ihren Übersichtsartikeln Grundsätze zur adäquaten Konformerenberechnung, gebräuchliche Programme zur Konformerengeneration sowie die zugrundeliegenden Algorithmen.

Kapitel 9

Motivation und Zielsetzung

Werden ligandbasiert Pharmakophor-Modelle erstellt, besteht eine der entscheidenden Schwierigkeiten darin, die Moleküle sinnvoll zu überlagern. Dies kann insbesondere bei sich stark unterscheidenden Grundgerüsten eine große Herausforderung darstellen. Da die Anzahl der frei verfügbaren Röntgenkristallstrukturen stetig zunimmt und in den letzten Jahren förmlich explodiert ist, bietet der strukturbasierte Pharmakophor-Ansatz eine vielversprechende Alternative, insbesondere dann, wenn bereits ein Ligand mit der Zielstruktur co-kristallisiert ist. In diesem Fall gibt die Orientierung des Liganden in der Bindetasche erste Hinweise darauf, welche Sub-Bindetaschen adressiert werden. Zusätzlich können auf der Basis von Ligand-Rezeptor-Interaktionen mutmaßlich wichtige Eigenschaften (interagierende Ligandeigenschaften) von vermeintlich unwichtigen Eigenschaften (nicht-interagierende Ligandeigenschaften) abgegrenzt und für die Ableitung strukturbasierter Pharmakophore genutzt werden.

Trotz wertvoller Hilfe, die viele Programme bei der automatisierten Interpretation der Ligand-Rezeptor-Interaktion leisten, um erste Anhaltspunkte für ein sinnvolles Modell zu erhalten, bleibt die Pharmakophor-Validierung und -Optimierung ein zeitraubender Prozess. Allein aufgrund der Tatsache, dass eine pharmakophore Ligandeigenschaft mit der Bindetasche interagiert, kann nicht auf deren Relevanz geschlossen werden⁹¹. Insbesondere wenn sehr viele Ligand-Rezeptor-Kontakte ausgebildet werden, ist es häufig sehr schwierig, essentielle Eigenschaften nach rationalen Kriterien aus der Fülle möglicher Kombinationen zu ermitteln.

Denkbare Lösungsansätze sind :

- Ist der katalytische Mechanismus der Zielstruktur bekannt, so können gezielt an der Reaktion beteiligte Ligand-Rezeptor-Interaktionen in der Pharmakophor-Hypothese berücksichtigt werden.
- Sind mehrere Kristallstrukturen der Zielstruktur vorhanden, können häufig auftretende Ligand-Rezeptor-Interaktionen Hinweise auf wichtige Schlüssel-Interaktionen geben (z.B. die Interaktion zur *Hinge*-Region bei nahezu allen Kinase-Inhibitoren).

Beide Ansätze erfordern ein recht hohes Maß an Information, die in vielen Fällen zu Beginn der Wirkstoffsuche jedoch nicht vorhanden ist.

Die in Kapitel 5 gezeigten Beispiele demonstrieren, wie die gezielte Verwendung und Interpretation von MIF-Informationen erfolgreich verwendet werden kann, um Leitstrukturen zu modifizieren, bzw. Pharmakophore auf Basis von Apo-Strukturen zu erstellen - auch wenn in diesen Fällen nur wenig Strukturinformation verfügbar ist. Jene Ansätze haben gemeinsam, dass sie individuelle Lösungsansätze für die jeweilige Zielstruktur darstellen. MIFs wurden individuell konturiert und pharmakophore Sphären nach visueller Inspektion in der Bindetasche platziert.

Die in dieser Arbeit vorgestellte Methode CavKA (Cavity Knowledge Acceleration) nimmt sich dieser Problematik an. Analog zu Ligandscout werden Ligand-Rezeptor-Komplexe automatisch nach geometrischen Regeln auf Ligand-Rezeptor-Interaktionen hin analysiert. Interagierende Ligandeigenschaften können mit zuvor berechneten GRID-MIFs abgeglichen werden, um eine Priorisierung pharmakophorer Ligandeigenschaften zu ermöglichen. Im Gegensatz zu den in Kapitel 5 genannten Beispielen wurden, basierend auf Vorversuchen, möglichst allgemein gültige Konturierungsgrade für jede Sonde definiert, die zur Visualisierung von *Hotspots* genutzt werden können.

Auch wenn sich bei H-Brücken-Donor- und -Akzeptor-Interaktionen der Betrag zur freien Energie nur sehr schlecht abschätzen oder vorhersagen lässt, so ist aufgrund ihrer gut beschriebenen Geometrie in den meisten Fällen recht eindeutig festzustellen, ob es zu einer Interaktion zwischen Ligand und Rezeptor kommt oder nicht (s. Abschn. 3.3). Bei einer hydrophoben Wechselwirkung ist dies nicht möglich (s. Abschn. 3.2). In diesen Fällen können MIFs hydrophober Sonden wertvolle Informationen über mutmaßlich energetisch günstige, hydrophobe Interaktionen liefern.

Die Entwicklung CavKAs hatte zum Ziel:

- Die automatisierte Identifikation von Ligand-Rezeptor-Interaktionen auf der Basis von Ligand-Rezeptor-Komplexen nach geometrischen Regeln
- Die zuverlässige Identifikation von Schlüsseigenschaften durch Kombination geometrischer Regeln und konturierter MIFs
- Die Identifikation wichtiger Wassermoleküle durch konturierte MIFs
- Die Erstellung von Pharmakophor-Modellen mit möglichst wenig Einfluss des Anwenders
- Die Entwicklung von Pharmakophor-Modellen mit höherer Erfolgsquote, verglichen mit rein geometrischen Ansätzen (Ligandscout)

Auch wenn Methoden zur Pharmakophor-Ableitung mittlerweile sehr benutzerfreundlich sind und gute Ansatzpunkte liefern, ist eine visuelle Inspektion der erstellten Modelle unabdingbar⁷⁶. CavKA stellt diesbezüglich keine Ausnahme dar. Es soll als ein komplementärer Ansatz geometrischer und energetischer Kriterien zur Ableitung von Pharmakophoren verstanden werden, die sich gegenseitig ergänzen, um erste Modellvorschläge zu liefern. CavKA kann und soll nicht den medizinischen Chemiker, die kritische Betrachtung des Modells und dessen erforderliche Validierung sowie die intensive Auseinandersetzung mit der zu untersuchenden Zielstruktur ersetzen.

Ein weiterer Kritikpunkt der in Kapitel 6 genannten Methoden ist deren Validierung. Während in der Ligandscout-Publikation nur die Methode selbst beschrieben wird¹⁵⁹, beschränkt sich die Validierung von Pocket 2.0 darauf, retrospektiv eine sehr begrenzte Anzahl von literaturbekannten Pharmakophoren zu reproduzieren¹⁶⁷. Die GBMP-Methode wurde auf nur zwei Zielstrukturen validiert¹⁶⁵, wobei die genauen Konturierungsgrade der GRID-MIFs nicht genannt wurden. Aufgrund dieser geringen Anzahl an unterschiedlichen Zielstrukturen besteht die Gefahr, dass die GBMP-Methode nicht universell einsetzbar ist. SALAM et al. verwendeten für die Validierung ihres E-Pharmakophor-Ansatzes eine deutlich größere Anzahl von Zielstrukturen (16 Zielstrukturen)¹⁶⁴. Jedoch wurden hierbei für alle Datensätze der aktiven Moleküle der gleiche Satz an Köder-Molekülen verwendet. Eine Validierung ist nur aussagekräftig, wenn Aktive und Köder-Moleküle sich nicht mit einfachen Mitteln voneinander separieren lassen. Da für alle Zielstrukturen die gleichen Köder-Moleküle verwendet worden sind, ist es möglich, dass diese Voraussetzung in

der Arbeit von SALAM et al.¹⁶⁴ nicht gegeben war. Es besteht die Gefahr, dass es sich bei den beobachteten Anreicherungen des E-Pharmakophor-Ansatzes um scheinbare Anreicherungen handelt, die letztlich auf ein mangelhaftes Datensatz-Design zurückzuführen sind und nicht auf die Leistungsfähigkeit der Methode selbst. Somit war neben der CavKA-Entwicklung die Validierung nach strengen Kriterien ein Kernaspekt dieser Arbeit:

- Die mit CavKA erstellten Pharmakophor-Modelle wurden mit Ligandscout-Modellen (jeweils Harte-Sphären-Modelle), dem Industriestandard für automatisierte Pharmakophor-Ableitungen, verglichen.
- Die CavKA-Gauß-Modelle wurden mit dem rein ligandzentrischen Ansatz Pharaos sowie anderen, weit verbreiteten ligandzentrischen Standardmethoden verglichen, um die Leistungsfähigkeit Pharaos bezüglich der Definition pharmakophorer Eigenschaften und der Superpositionierung realistisch einschätzen zu können^I.
- Vier der von SALAM et al. publizierten E-Pharmakophore¹⁶⁴ wurden für Datenbanksuchen auf dem in dieser Arbeit verwendeten Fieldscreen-Datensatz verwendet, um deren Leistungsfähigkeit mit den CavKA-Pharmakophoren für diese Zielstrukturen zu vergleichen^{II}.

^ICavKA verwendet die in Pharaos¹⁸⁷ verwendeten Regeln zur Definition pharmakophorer Ligandeigenschaften (s. Abschn. 10.1) .

^{II}Der von SALAM et al. verwendete Datensatz¹⁶⁴ und der hier verwendete Fieldscreen-Datensatz¹⁸⁴ besitzen nur vier gemeinsame Zielstrukturen.

Teil II

Methoden

Kapitel 10

CavKA: Implementierung pharmakophorer Eigenschaften

10.1 Pharmakophore Ligandeigenschaften

Definitionen von Ligandeigenschaften spielen eine entscheidende Rolle bei Pharmakophor-Suchen und können von Methode zu Methode stark variieren²¹⁴. SPITZER et al. erstellten auf Basis derselben Ligand-Rezeptor-Komplexe mit Phase, Catalyst und MOE scheinbar identische, strukturbasierte Pharmakophor-Modelle. Überraschenderweise erhielten die Autoren für jede Methode stark unterschiedlich zusammengesetzte Hitlisten²¹⁵. Diese Untersuchungen zeigen, wie sehr die Definition von Eigenschaften sowie die Superpositionierung von Referenz- und Datenbank-Molekül die Ergebnisse letztlich beeinflussen. Darüber hinaus verdeutlichen sie, dass ein Pharmakophor-Modell nicht ohne weiteres auf ein anderes Programm zur Datenbanksuche übertragen werden kann.

In CavKA wurden die in Pharaos genutzten Eigenschafts-Definitionen verwendet^{187,203}, die im Wesentlichen auf GREENE et al. zurückgehen¹⁸⁸. Dieses Vorgehen ermöglicht eine direkte Vergleichbarkeit des strukturbasierten CavKA-Ansatzes mit der ligandzentrischen Pharaos-Methode. Pharaos verwendet für die Definition der Eigenschaften keine SMARTS-Substruktursuchen²⁰⁶, sondern prüft den Molekülgraphen auf Atomtypen und deren Konnektivitäten über *Open Babel* Routinen^{216,217}. Eine umfassende Übersicht über SMARTS-Definitionen zur Identifikation pharmakophorer Ligandeigenschaften kann in der Arbeit von KOSSNER gefunden werden¹³.

Zusätzlich zu den Koordinaten der jeweiligen Eigenschaft definiert Pharao für H-Brücken-Donor-, H-Brücken-Akzeptor- und aromatische Eigenschaften projizierte Punkte in 1Å Entfernung von der jeweiligen Eigenschafts-Koordinate. Mit diesen projizierten Punkten lässt sich (optional) die Direktionalität dieser Eigenschaften im Rahmen der Datenbanksuche berücksichtigen. Die von Pharao verwendeten Regeln werden nachfolgend erläutert.

10.1.1 Aromatische Eigenschaften

Die Definition aromatischer Eigenschaften umfasst eine Ring- und eine Aromatizitäts-Detektion¹⁸⁷. Dazu wird zunächst die '*Kleinste Menge an kleinsten Ringen*' (engl. *smallest set of smallest rings*) identifiziert²¹⁸. Ein Ring wird als aromatische Eigenschaft definiert, sofern er planar ist, keine exocyclischen Doppelbindungen besitzt und die HÜCKEL-Regel erfüllt, nach der ein aromatisches System $4n+2\pi$ Elektronen besitzen muss²¹⁹.

CavKA definiert zusätzlich zu den Koordinaten des aromatischen Ring-Zentrums jeweils einen Punkt ober- und unterhalb orthogonal zur Ringebene in 2.1Å Entfernung. Diese projizierten Punkte werden benötigt um $\pi - \pi$ -Ligand-Rezeptor-Interaktionen berechnen zu können (vgl. auch Abschn. 11.5).

10.1.2 Hydrophobe Eigenschaften

Die Definition hydrophober Ligandeigenschaften wird viel diskutiert und variiert zum Teil stark^{76,220}, weshalb je nach Methode die Definition sehr unterschiedlich ausfallen kann^{214,215}.

Pharao¹⁸⁷ greift auf einen inkrementellen Ansatz von GREENE et al. zurück¹⁸⁸. Hierbei wird regelbasiert ein Hydrophobizitäts-Wert h für Ligandeigenschaften definiert:

$$h = t \cdot s \tag{10.1}$$

wobei t den topologieabhängigen Lipophilie-Faktor darstellt, der von der Art des Atoms abhängt (vgl. Tab. 10.1) und s die Solvens zugängliche Fläche (engl. *solvent accessible surface area*; Abk. SASA)¹⁸⁸.

Um redundante Information zu reduzieren, werden die einzelnen Atome mit den zugeordneten Lipophilie-Werten in Gruppen zusammengefasst.

Eine Gruppe wird wie folgt definiert¹⁸⁷:

- ein Ring mit sieben oder weniger Mitgliedern,
- Atome mit drei oder mehreren Bindungen zusammen mit ihren Nachbaratomen (keine Wasserstoffatome),
- die verbleibenden Atome werden anhand ihrer Konnektivität als Kette betrachtet, die ihrerseits eine Gruppe bilden. Ringe mit mehr als sieben Mitgliedern werden ebenfalls als Gruppe definiert.

Für jede der auf diese Weise definierten Gruppen werden die h -Werte aufsummiert. Eine Ligandeigenschaft wird als hydrophob definiert, wenn ein Grenzwert ($h_{min} > 9.87$)¹ überschritten wird¹⁸⁷. Aromatische Ringe werden somit nicht zwangsweise als hydrophob definiert, sondern nur wenn der Hydrophobizitäts-Grenzwert überschritten wird. Eine detaillierte Beschreibung des Algorithmus zur Definition hydrophober Eigenschaften geben GREENE et al.¹⁸⁸.

Ähnliche Algorithmen, die die Hydrophobizität inkrementell berechnen, sind von WILDMANN und CRIPPEN verwendet worden um logP-Werte vorherzusagen²²¹.

Tabelle 10.1: Definitionen des topologieabhängigen Lipophilie-Faktors t sowie Beiträge zur Lipophilie (nach GREENE et al.¹⁸⁸).

Kategorie	t	Beschreibung
1	0	N, O oder H
2	0	S in SH
3	0	≤ 2 Bindungen entfernt von einem geladenen Atom
4	0	≤ 2 Bindungen entfernt von OH oder NH ohne delokalisierte Elektronen
5	0	≤ 1 Bindung entfernt von SH ohne delokalisierte Elektronen
6	0	≤ 2 Bindungen entfernt von O mit einer Doppelbindung
7	0	≤ 1 Bindung entfernt von S mit einer Valenz > 2
8	0	S mit Doppelbindung
9	0.6	3 Bindungen entfernt von einem O mit Doppelbindung
10	0.6	2 Bindungen entfernt von S mit einer Valenz > 2
11	0.6	1 Bindung entfernt von einem S mit Doppelbindung
12	0	2 oder mehr erfüllte Bedingungen der Kategorien 9-11
13	0.25	1 benachbartes O oder N ohne delokalisierte Elektronen
14	0	> 1 benachbartes O oder N ohne delokalisierte Elektronen
15	1	zu keiner der Kategorien 1-14 gehörend

¹Dies entspricht der Hälfte des h -Wertes einer endständigen, exponierten Methyl-Gruppe^{187,188}.

10.1.3 H-Brücken-Donor-Eigenschaften

Als H-Brücken-Donoren werden in Pharao definiert¹⁸⁷:

- Atome, die Stickstoff oder Sauerstoff sind,
- die keine negative Formalladung besitzen,
- die mindestens ein kovalent gebundenes Wasserstoffatom besitzen.

Zusätzlich zu den Koordinaten der H-Brücken-Donor-Eigenschaft (Abk. Don) werden im Rahmen der CavKA-Methode die Koordinaten der gebundenen Wasserstoffatome definiert. Über den so definierten Vektor \vec{DonH} werden im Folgenden Winkelberechnungen durchgeführt.

10.1.4 H-Brücken-Akzeptor-Eigenschaften

Als H-Brücken-Akzeptoren werden in Pharao definiert¹⁸⁷:

- Atome, die Stickstoff sind,
 - wenn sich der Stickstoff in einem aromatischen Ringsystem befindet, muss er weniger als drei Bindungen aufweisen,
 - ausgenommen ist Stickstoff aus Sulfonamiden oder Amiden,
 - ausgenommen ist an Aromaten gebundener Stickstoff mit drei Bindungen (Anilin-Stickstoff).
- oder Sauerstoff sind,
- die keine positive Formalladung besitzen,
- die mindestens ein freies Elektronenpaar besitzen,
- die sterisch in der Lage sind eine H-Brückenbindung einzugehen (engl. *accessible*).

Um das letztgenannte Kriterium zu prüfen, werden um ein mutmaßliches H-Brücken-Akzeptor-Atom (Abk. Acc) im Abstand von 1.8Å gleichmäßig verteilte Punkte projiziert, die Wasserstoffatome darstellen sollen, mit denen der potentielle H-Brücken-Akzeptor interagieren kann. Eine Akzeptor-Eigenschaft wird definiert, wenn mindestens 2% der Punkte zugänglich sind, d.h. sie nicht mit Nachbaratomen kollidieren. In der CavKA-Methode wird neben den Koordinaten des H-Brücken-Akzeptors ein beliebiges (schweres) Nachbaratom definiert (Abk. X). Über den so definierten Vektor \vec{XAcc} lassen sich im Folgenden Winkelberechnungen durchführen.

10.1.5 Ionische Eigenschaften

Ionische Eigenschaften werden nur anhand einer vorhandenen Formalladung definiert und am Ort des entsprechenden Atoms platziert. Bei einer negativen Formalladung wird eine anionische Eigenschaft, bei positiven Formalladungen eine kationische Eigenschaft definiert¹⁸⁷. Protonierungszustände des Referenz-Liganden (und hieraus resultierende Formalladungen) werden über die *Protonate-3D*-Methode im Kontext der Bindetasche definiert²²².

Nachdem pharmakophore Eigenschaften sowie die zugehörigen Koordinaten definiert worden sind, bilden diese den Ausgangspunkt für weitere Berechnungen. Die zusätzlich bei den gerichteten Interaktionen (H-Brücken-Donor, -Akzeptor, $\pi - \pi$ -Interaktionen) definierten Punkte sind erforderlich, um Winkelberechnungen im Rahmen der geometrischen Ligand-Rezeptor-Interaktionserkennung (s. Kap. 11) zu realisieren.

10.2 Pharmakophore Proteineigenschaften

Die Zuweisung pharmakophorer Eigenschaften an *Aminosäureatome* erfolgt regelbasiert (modifiziert nach DREHER^{II})²²³. Aminosäureatome werden unter Zuhilfenahme einer Liste oder eines Wörterbuches pharmakophore Eigenschaften zugeordnet. Listen werden zur Definition nicht-gerichteter Proteineigenschaften (wie z.B. hydrophobe, kationische und anionische Aminosäureeigenschaften) verwendet. Im Gegensatz dazu werden Wörterbücher zur Definition gerichteter Proteineigenschaften (H-Brücken-Donor- und -Akzeptor-Eigenschaften) genutzt. Eine Ausnahme stellen diesbezüglich aromatische Aminosäureeigenschaften dar ($\pi - \pi$ -Interaktionen, bzw. Kationen- π -Interaktionen), die gerichteter Natur sind und über verkettete Listen definiert werden.

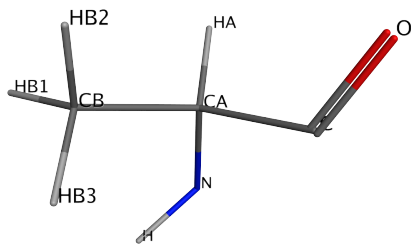
^{II}Dreher definierte in seinem Kinase Interaktions Fingerprint (Abk. KIFT) aromatische Aminosäureatome nicht explizit. Im KIFT werden keine $\pi - \pi$ -Interaktionen identifiziert, sondern nur aromatisch-hydrophobe Interaktionen.

Die makromolekulare Information der Röntgenkristallstruktur muss im .pdb Format vorliegen und wird mit Biopython^{224,225} unter Verwendung des Moduls *Bio.PDB*²²⁶ eingelesen. Das so erhaltene Biopython-Strukturobjekt ist wie folgt absteigend hierarchisch gegliedert²²⁷:

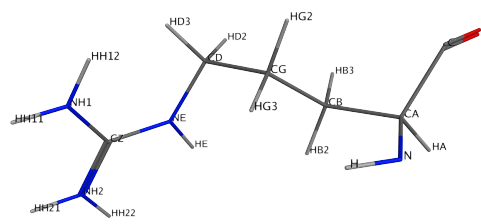
1. Struktur-Objekt (engl. *structure*)
2. Modell (engl. *model*)
3. Aminosäuren-Kette (engl. *chain*)
4. Aminosäure-Rest (engl. *residue*)
5. Aminosäuren-Atom (engl. *atom*)

Das Biopython-Objekt kann iterativ prozessiert werden, sodass z.B. die zu einem Aminosäure-Rest zugehörigen Atome ausgegeben werden können. Im Nachfolgenden wird überwiegend auf die pharmakophoren Eigenschaften der Aminosäureatome eingegangen. So steht die Abkürzung 'LEU_CB' für das β -C-Atom (Abk. 'CB'; Bio.PDB-Nomenklatur) des Aminosäure-Rests Leucin (Abk. 'LEU'; Bio.PDB-Nomenklatur). Da sich 'LEU_CB' in der Liste mit hydrophoben Eigenschaften befindet, werden diesem Atom hydrophobe Eigenschaften zugewiesen. Zusätzlich können auf der Atomebene die dazugehörigen Koordinaten über die *Bio.PDB*-Methode '*get_coord()*' ausgegeben werden, was für Distanz- und Winkelberechnungen erforderlich ist. Abbildung 10.1 gibt eine Übersicht über die Syntax der Atomnamen der 20 Aminosäuren.

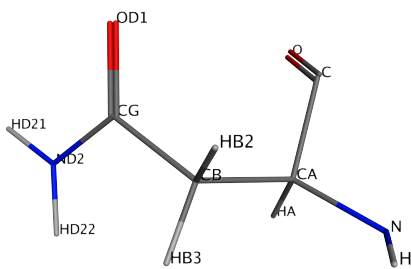
Die MOE-PDB-Nomenklatur weicht bei der Bezeichnung von vierstelligen Atomnamen (nur bei Wasserstoffen vorhanden) minimal von der in Bio.PDB verwendeten ab. So lautet in MOE beispielsweise die Bezeichnung für eines der H-Atome 'HH21', das an das N-Atom 'NH2' im Arginin gebunden ist. In Bio.PDB wird zuerst die Nummer des H-Atoms (1), dann die Atom-Bezeichnung (HH) und schließlich die Nummer des Rests genannt, an den das H-Atom gebunden ist (2). Die vollständige Bezeichnung des Atoms in Bio.PDB würde also '1HH2' statt 'HH21' lauten.



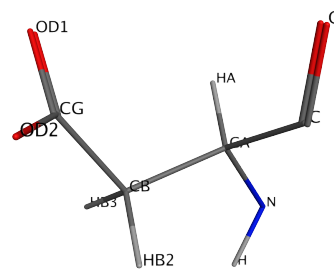
(a) Alanin (ALA)



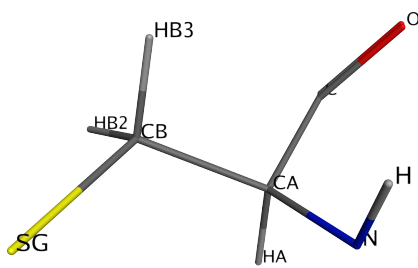
(b) Arginin (ARG)



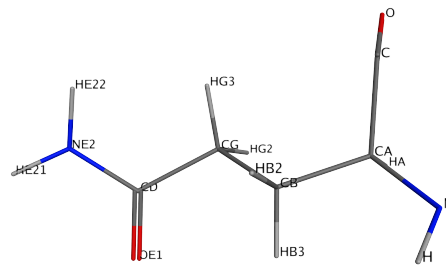
(c) Asparagin (ASN)



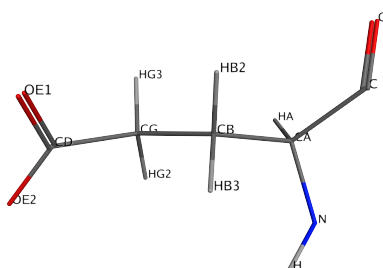
(d) Asparaginsäure (ASP)



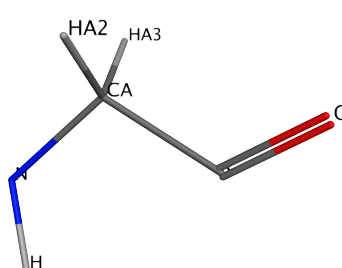
(e) Cystein (CYS)



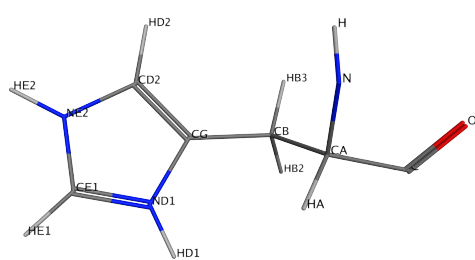
(f) Glutamin (GLN)



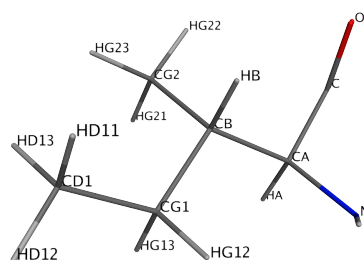
(g) Glutaminsäure (GLU)



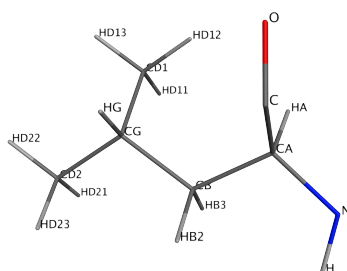
(h) Glycin (GLY)



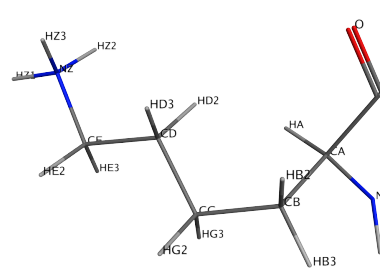
(i) Histidin (HIS)



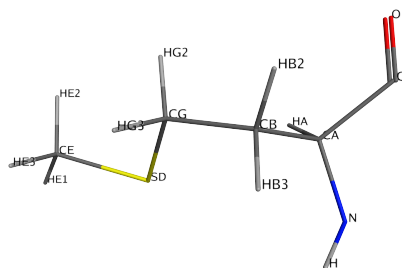
(j) Isoleucin (ILE)



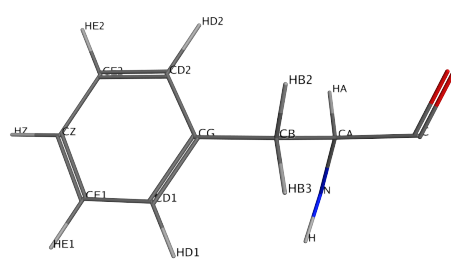
(k) Leucin (LEU)



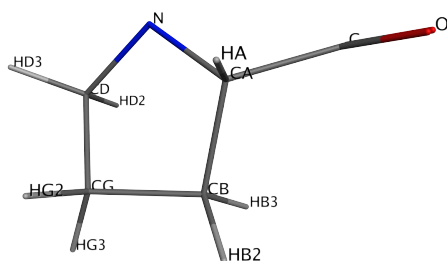
(1) Lysin (Lys)



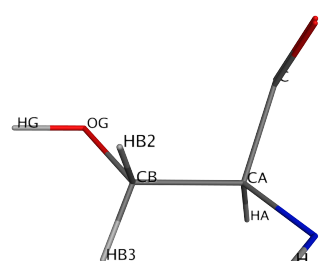
(m) Methionin (MET)



(n) Phenylalanin (PHE)



(o) Prolin (PRO)



(p) Serin (SER)

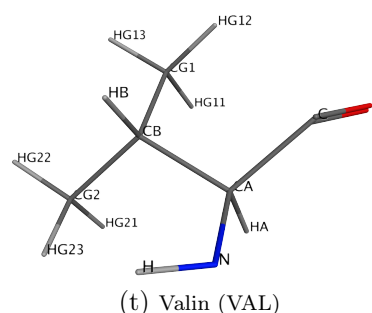
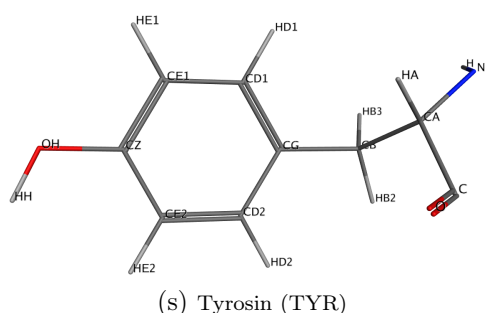
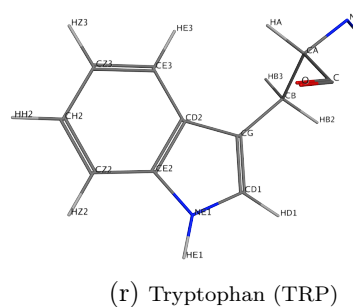
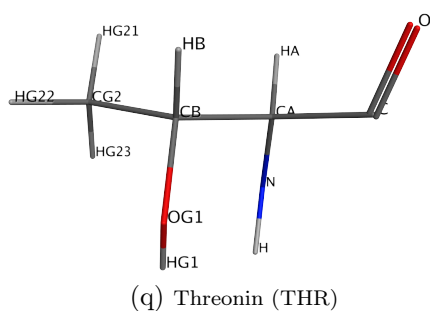


Abbildung 10.1: Übersicht über Atomnamen der einzelnen Aminosäuren. Hierzu wurden exemplarisch aus der Röntgenkristallstruktur PDB 1o86 (*Angiotensin-Konversionsenzym*) für jede der 20 Aminosäuren (Abkürzung in Klammern) ein beliebiger Vertreter gewählt und isoliert mit den Atomnamen in MOE⁶² dargestellt. Die Hydroxylfunktion der jeweiligen Aminosäure, die eine Peptidbindung eingeht, ist nicht dargestellt.

10.2.1 Listen pharmakophorer Proteineigenschaften

Ionische und hydrophobe Eigenschaften werden über Listen definiert. Da bei diesen Eigenschaften Ligand-Rezeptor-Interaktionen ausschließlich distanzbasiert ermittelt werden, wird auf die Definition weiterer Punkte (wie sie für Winkelberechnungen benötigt werden) verzichtet.

```
Hydrophobe_AS_Eigenschaften = ['LEU_CB', 'LEU_CG', 'LEU_CD1', 'LEU_CD2', 'PRO_CB',
'PRO_CG', 'PRO_CD', 'TYR_CB', 'TYR_CG', 'TYR_CD1', 'TYR_CD2', 'TYR_CE1', 'TYR_CE2',
'TYR_CZ', 'LYS_CB', 'LYS_CG', 'LYS_CD', 'LYS_CE', 'TRP_CB', 'TRP_CG', 'TRP_CD1', 'TRP_CD2',
'TRP_CE2', 'TRP_CE3', 'TRP_CZ2', 'TRP_CZ3', 'TRP_CH2', 'PHE_CB', 'PHE_CG', 'PHE_CD1',
'PHE_CD2', 'PHE_CE1', 'PHE_CE2', 'PHE_CZ', 'ARG_CB', 'ARG_CG', 'ARG_CD', 'VAL_CB',
'VAL_CG1', 'VAL_CG2', 'ILE_CB', 'ILE_CG1', 'ILE_CG2', 'ILE_CD1', 'MET_CB', 'MET_CG', 'MET_SD',
'MET_CE']
```

```
Kationische_AS_Eigenschaften = ['LYS_NZ', 'ARG_NE', 'HIS_NE2']
```

```
Anionische_AS_Eigenschaften = ['GLU_OE1', 'GLU_OE2', 'GLU_CD', 'ASP_CG', 'ASP_OD1',
'ASP_OD2']
```

10.2.2 Wörterbücher pharmakophorer Proteineigenschaften

Bei den Atomen, die die Eigenschaft H-Brücken-Donor oder -Akzeptor codieren, werden zusätzlich über ein Wörterbuch (engl. *dictionary*; ein Python Objekt) assoziierte Nachbaratome definiert. Die zugewiesenen Nachbaratome (z.B. kovalent an einen H-Brücken-Donor gebundene Wasserstoffatome, bzw. ein schweres Nachbaratom bei einem H-Brücken-Akzeptor) sind nötig, um die Koordinaten der jeweiligen Atome abfragen zu können und die in Tabelle 11.1 aufgeführten Winkelkriterien abprüfen zu können.

So ist der Wörterbucheintrag 'GLU_OE1':['CD'] wie folgt zu deuten: Bei 'GLU_OE1' handelt es sich um das Atom 'OE1' eines Glutamat-Rests 'GLU', das H-Brücken-Akzeptor-Eigenschaften besitzt (es ist in dem Wörterbuch mit **H-Brücken-Akzeptor_AS_Eigenschaften** definiert). Mit ':['CD']' ist das Nachbar-C-Atom definiert, das an 'GLU_OE1' gebunden ist. Alle weiteren Einträge sind analog zu interpretieren.

```
H-Brücken-Akzeptor_AS_Eigenschaften =
```

```
### Seitenkette
```

```
{'TYR_OH':['CZ'], 'SER_OG':['CB'], 'THR_OG1':['CB'], 'ASP_OD1':['CG'], 'GLU_OE1':['CD'],
'GLN_OE1':['CD'], 'ASN_OD1':['CG'], 'CYS_SG':['CB'],
```

```
### Hauptkette
```

```
'ALA_O':['C'], 'ARG_O':['C'], 'ASN_O':['C'], 'ASP_O':['C'], 'CYS_O':['C'], 'GLN_O':['C'], 'GLU_O':['C'],
'GLY_O':['C'], 'HIS_O':['C'], 'ILE_O':['C'], 'LEU_O':['C'], 'LYS_O':['C'], 'MET_O':['C'], 'PHE_O':['C'],
'PRO_O':['C'], 'SER_O':['C'], 'THR_O':['C'], 'TRP_O':['C'], 'TYR_O':['C'], 'VAL_O':['C'], 'ASP_OD2':['CG'],
'GLU_OE2':['CD']}
```

```
H-Brücken-Donor_AS_Eigenschaften =
```

```
### Seitenkette
```

```
{'ARG_NE':['HE'], 'ARG_NH1':['1HH1','2HH1'], 'ARG_NH2':['1HH2','2HH2'], 'ASN_ND2':['1HD2','2HD2'],
'GLN_NE2':['1HE2','2HE2'], 'HIS_NE2':['2HE'], 'LYS_NZ':['1HZ','2HZ','3HZ'], 'SER_OG':['HG'],
'THR_OG1':['1HG'], 'TRP_NE1':['1HE'], 'TYR_OH':['HH'],
```

```
### Hauptkette
```

```
'ALA_N':['H'], 'ARG_N':['H'], 'ASN_N':['H'], 'ASP_N':['H'], 'CYS_N':['H'], 'GLN_N':['H'], 'GLU_N':['H'],
'GLY_N':['H'], 'HIS_N':['H'], 'ILE_N':['H'], 'LEU_N':['H'], 'LYS_N':['H'], 'MET_N':['H'], 'PHE_N':['H'],
'PRO_N':['H'], 'SER_N':['H'], 'THR_N':['H'], 'TRP_N':['H'], 'TYR_N':['H'], 'VAL_N':['H']}
```

10.2.3 Verkettete Listen aromatischer Proteineigenschaften

Bei den gerichteten aromatischen Interaktion werden aromatische Aminosäure-Reste zunächst auf *Reste-Ebene* identifiziert:

Aromatische_AS_Eigenschaften (Aminosäure-Rest-Ebene)

```
AROMATIC = ['PHE','TRP', 'HIS', 'TYR']
```

Alle Aminosäureatome des jeweiligen Rests werden iterativ abgearbeitet und anhand weiterer Listen wird identifiziert, welche Atome zum jeweiligen aromatischen Ring gehören.

Aromatische_AS_Eigenschaften (Aminosäureatom-Ebene)

```
TRP_local_I = ['CE2', 'CD2', 'CE3', 'CZ3', 'CH2', 'CZ2']
```

```
TRP_local_II = ['CE2', 'CD2', 'NE1', 'CD1', 'CG']
```

```
PHE_local    = ['Phe_CG', 'Phe_CZ', 'Phe_CD1', 'Phe_CD2', 'Phe_CE1', 'Phe_CE2']
```

```
TYR_local    = ['Tyr_CG', 'Tyr_CZ', 'Tyr_CD1', 'Tyr_CD2', 'Tyr_CE1', 'Tyr_CE2']
```

```
HIS_local    = ['His_CE1', 'His_CD2', 'His_CG', 'His_ND1', 'His_NE2']
```

Von diesen aromatischen Ringatomen werden die Koordinaten gespeichert und ausgehend hiervon das Ring-Zentrum und die projizierten Punkte ober- und unterhalb der Ringebene berechnet.

10.3 Pharmakophore Modelleigenschaften

Im Vorherigen wurde erläutert, wie pharmakophore Eigenschaften auf Ligand-, bzw. Proteinseite definiert werden. In diesem Abschnitt wird darauf eingegangen, wie die Definition pharmakophorer Eigenschaften in *Pharmakophor-Hypothesen* erfolgt, mit denen letztendlich eine Datenbank durchsucht wird. Mit CavKA als Schlüssel-Eigenschaft identifizierte Ligandeigenschaften können für die Datenbanksuche in ein Pharmakophor-Modell exportiert werden. Derzeit werden die Dateiformate .phar (Pharao) und .ph4 (MOE) unterstützt.

Eine Übersicht über pharmakophore Modelleigenschaften sowie zugehörige Parameter wird in Tabelle 10.2 gegeben. Die Pharao- und MOE-Pharmakophor-Modelle werden nachfolgend detaillierter erläutert. Pharao bietet die Möglichkeit, gerichtete pharmakophore Eigenschaften zu definieren. Hierbei wird zusätzlich zu der Koordinate der jeweiligen Eigenschaft (Abk. P) eine weitere Koordinate (ein sogenannter projizierter Punkt; Abk. Q) definiert. Der Vektor \vec{PQ} definiert die *Direktionalität* einer Eigenschaft (z.B. entlang der C=O Bindung bei einer Carbonylfunktion).

Die Direktionalität solcher Eigenschaften kann optional bei einer Datenbanksuche berücksichtigt werden^{III}. Einen Überblick, wie die Direktionalität dieser gerichteten Eigenschaften in die Pharaos-Scoring-Funktion integriert ist, geben DE TAMINAU et al.¹⁸⁷.

Bei der Modell-Erstellung wurden keine Hybrid-Sphären (mehrere Eigenschaften für eine Sphäre) definiert. Jede Eigenschaft ist durch eine einzelne Sphäre berücksichtigt worden: So wird ein Phenylring nicht durch eine Sphäre mit den Eigenschaften 'hydrophob' und 'aromatisch' repräsentiert, sondern durch zwei separate Sphären mit der jeweiligen Eigenschaft.

Dieses Vorgehen ist gewählt worden, um die Spezifität von Pharmakophor-Modellen zu erhöhen. Hybrid-Sphären können, sofern gewünscht, manuell vom Anwender definiert werden. Bei Hybrid-Eigenschaften werden im Gegensatz zu den konventionellen Eigenschaften (mit Ausnahme ionischer und hydrophober Eigenschaften) keine Richtungsvektoren definiert. In Übereinstimmung mit KOSSNER, der für seinen ligandenzentrischen DIP²-Deskriptor analog projizierte Punkte (engl. *projected points*)¹³ verwendete, bestätigte sich in Vorversuchen, dass durch die Verwendung von Richtungsvektoren, die die Direktionalität einer Eigenschaft berücksichtigen, in vielen Fällen bessere Ergebnisse erzielt werden können.

Tabelle 10.2: Übersicht pharmakophorer Modelleigenschaften in Pharaos und MOE. Weitere Erläuterungen siehe Text. (*P*: *pharmakophore Eigenschaft*; σ : *Breite des dreidimensionalen Gauß-Peaks in Å*; *r*: *Radius der Sphäre in Å*).

Eigenschaft	Pharaos (Gauß-Modell)			MOE (Hartes-Sphären-Modell)	
	<i>P</i> (Pharaos Code)	gerichtete Eigenschaft	σ (Å)	<i>P</i> (MOE Code)	<i>r</i> (Å)
aromatischer Ring	ARO	ja	0.7	Aro	1.4
hydrophob	HYD	nein	1.0	Lipo	1.4
aromatisch & hydrophob	HYBL	nein	0.7	-	-
H-Brücken-Donor	HDON	ja	1.0	Don	1.0
H-Brücken-Akzeptor	HACC	ja	1.0	Acc	1.0
H-Brücken-Donor & H-Brücken-Akzeptor	HYBH	nein	1.0	-	-
positiv geladen	POSC	nein	1.0	Cat	1.4
negativ geladen	NEGC	nein	1.0	Ani	1.4
Excluded Volume	EXCL	nein	1.7	Excl	1.0

^{III}Richtungsvektoren in einem Pharmakophor-Modell sind nicht zu verwechseln mit den zusätzlich definierten Punkten, z.B. von Protein-Seite oder Ligand-Seite aus, die für die Interaktionserkennung zwischen Ligand und Rezeptor verwendet werden (vgl. Abschn. 11).

10.3.1 MOE-Pharmakophore (.ph4)

Bei den MOE-Pharmakophor-Modellen handelt es sich um ein Hartes-Sphären-Modell, das durch die jeweilige Eigenschaft p an einem Punkt P (P_x, P_y, P_z) im Raum definiert wird. Der Tolreanzradius r gibt hierbei an, wie weit ein Datenbank-Molekül mit einer kompatiblen Eigenschaft von X entfernt sein darf, damit der Pharmakophor dennoch erfüllt ist.

10.3.2 Pharao-Pharmakophore (.phar)

Bei den Pharao-Pharmakophor-Modellen handelt es sich um Gauß-Modelle, die durch die Koordinate eines Punktes P (P_x, P_y, P_z) im Raum mit der Eigenschaft p sowie der Breite σ (engl. *spread*) des dreidimensionalen Gauß-Peaks definiert werden. Bei gerichteten Eigenschaften (s. Tab. 10.2) ist zusätzlich in 1Å Entfernung von x eine zweite Koordinate Q (Q_x, Q_y, Q_z) definiert. Der Vektor \vec{PQ} definiert die Direktionalität einer Eigenschaft im Rahmen der Datenbanksuche.

10.3.3 Verbotene Sphären

Ausgehend von jedem Ligand-Atom sind im Umkreis von 6Å vollständig alle schweren Rezeptor-Atome als verbotene Sphären (engl. *excluded volume*) mit einem Durchmesser von 1Å definiert worden, in denen kein Datenbank-Molekül zum Liegen kommen darf.

Die vollständige Repräsentation eines Rezeptors durch verbotene Sphären wird kontrovers diskutiert⁷⁶: Ist es auf der einen Seite möglich, sehr spezifische Modelle zu erstellen, so birgt ein derartiges Vorgehen die Gefahr, dass sich die *falsch-negativ Rate* unnötig erhöht. Auf der anderen Seite kann sich die *falsch-positiv Rate* erhöhen, wenn nur Teile des Rezeptors als verbotene Bereiche definiert werden und Köder-Moleküle den Pharmakophor zwar formal erfüllen, jedoch durch 'Löcher' in den verbotenen Sphären penetrieren und mit schweren Atomen des Rezeptors kollidieren würden.

Kapitel 11

CavKA: Implementierung der geometrischen Ligand-Rezeptor-Komplex-Analyse

In diesem Kapitel wird erläutert, wie die zuvor definierten Ligand- und Proteineigenschaften in Ligand-Rezeptor-Komplexen mittels CavKA nach geometrischen Kriterien analysiert und auf Interaktionen hin untersucht werden. Tabelle 11.1 zeigt eine Übersicht der in CavKA abgeprüften Interaktionstypen sowie die dazugehörigen geometrischen Parameter, die nachfolgend erläutert werden.

Für Distanzen und Winkel sind dabei gezielt große Toleranzbereiche definiert worden. In Übereinstimmung mit BOEHM zeigte sich in Vorversuchen, dass bei zu strikten Vorgaben wichtige Schlüssel-Interaktionen nicht mehr erkannt werden¹⁷⁶. Zudem werden auch in Röntgenkristallstrukturen bei Ligand-Rezeptor-Interaktion häufig nicht-optimale Geometrien für H-Brücken-Bindungen beobachtet¹⁷⁶.

11.1 Kompatible Ligand-Protein Eigenschafts-Kombinationen

Bevor Winkelkriterien geprüft werden, um zu ermitteln, ob es zu einer Interaktion zwischen Ligand und Bindetasche kommt, wird die Kompatibilität von Ligand- und Rezeptoreigenschaften abgeglichen. Erlaubte Kombinationen (z.B. zwischen hydrophoben Eigenschaften oder zwischen einem H-Brücken-Donor und -Akzeptor) sind in Tabelle 11.1 unter *Interaktionstypen* aufgeführt.

Von jeder zuvor definierten Ligandeigenschaft werden über Pybel²²⁸ die entsprechenden Koordinaten abgefragt. Ausgehend hiervon wird mittels KD-Tree²²⁹ eine Nachbarschaftssuche durchgeführt¹, bei der innerhalb eines maximalen Suchradius nach kompatiblen Proteineigenschaften gesucht wird.

Sind innerhalb der definierten Maximaldistanz erlaubte Eigenschaftskombinationen gefunden worden, wird im nächsten Schritt geprüft, ob diese einen Mindestabstand einhalten. Dazu wird zwischen Ligand-Koordinaten $(x_{Lig}(i), y_{Lig}(i), z_{Lig}(i))$ und Protein-Koordinaten $(x_{Prot}(j), y_{Prot}(j), z_{Prot}(j))$, die über das Modul Bio.PDB²²⁶ ermittelt worden sind, die Euklidische Distanz $Dist_{Euklid}$ berechnet:

$$Dist_{Euklid}(i, j) = \sqrt{(x_{Lig}(i) - x_{Prot}(j))^2 + (y_{Lig}(i) - y_{Prot}(j))^2 + (z_{Lig}(i) - z_{Prot}(j))^2} \quad (11.1)$$

Ist die erforderliche Mindestdistanz überschritten, werden weitere Kriterien (wie z.B. Winkelkriterien, bzw. eine Mindestanzahl von Interaktionen) überprüft.

Eine Ausnahme bilden hierbei **Met**...**Ani** / **Acc** und **Kat**...**Ani** Interaktionen, die ausschließlich distanzbasiert und ohne weitere Nebenbedingungen ermittelt werden.

Tabelle 11.1: Übersicht der in CavKA (fett) verwendeten und zum Teil angepassten geometrischen Regeln zur Interaktionserkennung. In Klammern sind die in den Originalpublikationen genannten Werte aufgeführt. In der Spalte 'min. Interaktionen' ist aufgeführt, wie oft ein Interaktionstyp zustandekommen muss, damit eine Ligandeigenschaft als 'interagierend' definiert wird. Erläuterungen zu den Winkeln α, β, γ und δ s. Text.

Interaktionstypen	Distanz (Å) (Heav...Heav)	min. Interaktionen	α	$\beta/\gamma/\delta$
Hyd...Hyd	3.0-6.0	3 ²²³	-	-
$\pi_{Hyd} \dots \text{Hyd}$	(3.0-4.0 - Ali) ¹⁷³ (3.0-6.0 - π) ¹⁷³	1	-	-
Don...Acc	2.3-3.5	1	$110^\circ \leq \alpha \leq 180^\circ$ ¹⁷⁶	$90^\circ \leq \beta \leq 170^\circ$ ($150^\circ < \beta < 180^\circ$) ¹⁷⁶
Kat...Ani	2.3-4.5	1	-	-
Kat...π	2.5-4.5 (3.4-4.0) ¹⁰⁶	1	$125^\circ \leq \alpha \leq 180^\circ$	-
$\pi \dots \pi$ ('Face-to-Face')	3.0-6.0	1	$120^\circ \leq \alpha \leq 180^\circ$	$\beta - \alpha < \delta $; $ \delta = 15^\circ$
$\pi \dots \pi$ ('Edge-to-Face')	3.0-6.0	1		$140^\circ \leq \beta \leq 180^\circ$ $80^\circ \leq \gamma \leq 100^\circ$
Met...Ani / Acc	1.5-4.5 (opt: 2.0) ¹⁷⁶	1	-	-

Abkürzungen: Ali: aliphatisch; π : aromatisch; Hyd: hydrophob; π_{Hyd} : aromatisch-hydrophob; $\pi \dots \pi$: Pi-Pi; Don: H-Brücken-Donor; Acc: H-Brücken-Akzeptor; Kat: kationisch; Ani: anionisch; Heav: schweres Atom; Met: Metall-Ion; opt: optimaler Wert.

¹Es wurde die KD-Tree Implementierung in Biopython^{224,225} verwendet.

11.2 CavKA: Hydrophobe Interaktionen

Hydrophobe Interaktionen sind der dominierende Faktor für die Stärke von Ligand-Rezeptor-Interaktionen. Sie beruhen im Wesentlichen auf der Verdrängung von Wassermolekülen, sind ungerichteter Natur und gehen mit einer Erhöhung der Entropie einher. Im Gegensatz zu lokal fixierten Wassermolekülen lässt sich die Position normalbeweglicher Wassermoleküle *nicht* exakt durch kristallographische Analysen bestimmen. Die Effekte des Liganden auf das Wassernetzwerk der solvatisierten Bindetasche lassen sich daher nicht mit geometrischen Kriterien vorhersagen, weswegen bewusst eine sehr einfache Definition für hydrophobe Interaktionen gefasst worden ist.

Angelehnt an BOEHM¹⁷⁶ und DREHER²²³ werden hydrophobe, bzw. aromatisch-hydrophobe Ligandeigenschaften als interagierend erkannt, wenn sich hierzu mindestens drei hydrophobe Aminosäureatome²²³ in einer Distanz von 3-6Å befinden. Abbildung 11.1 illustriert die geometrischen Verhältnisse. Alternativ zu diesem Ansatz verwendete BOEHM das Maß an vergrabener lipophiler Oberfläche um lipophile Interaktionen zu bewerten²³⁰.

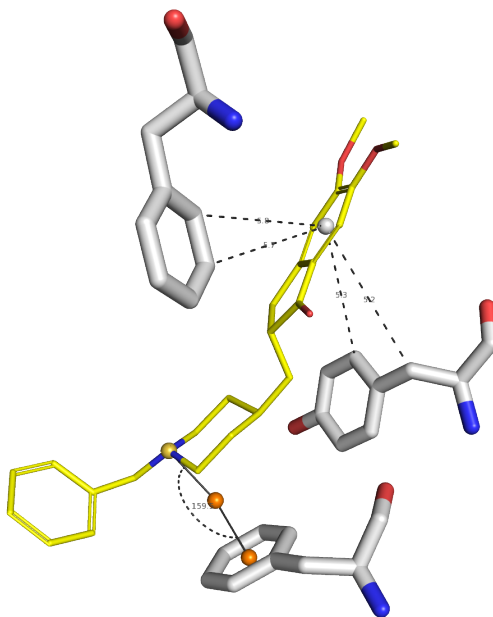


Abbildung 11.1: Hydrophobe- und Kationen- π -Ligand-Rezeptor-Interaktion. Für eine hydrophobe Interaktion (graue Sphäre) müssen mindestens drei Kontakte zu hydrophoben Aminosäureatomen in einer Distanz von 3-6Å ausgebildet werden (hier: vier Kontakte; gestrichelte Linien). Für eine Kationen- π -Interaktion müssen sich aromatisches und kationisches Zentrum in einer Distanz von 2.5-4.5Å befinden. Der Winkel α zwischen dem aromatischen Aminosäurezentrum (orange Sphäre), projiziertem Aminosäure-Punkt (orange Sphäre) und kationischer Eigenschaft (gelbe Sphäre) muss 125-180° betragen (hier: 159.5°). Weitere Erläuterungen s. Text. *Gelb: Ligand (Donepezil); grau: Protein (Acetylcholinesterase; PDB 1eve).*

11.3 CavKA: H-Brücken-Interaktionen

Bei H-Brücken-Bindungen handelt es sich um gerichtete Interaktionen, die für die Spezifität einer Ligand-Rezeptor-Interaktion verantwortlich sind. Im Gegensatz zur NMR-Spektroskopie sind Wasserstoffatome mittels Röntgendiffraktometrie nicht auflösbar, wobei Röntgenkristallstrukturen einen deutlich größeren Anteil in der Protein-Datenbank^{160,161} als Hauptressource für freiverfügbare Kristallstrukturen ausmachen¹.

Nicht aufgelöste Wasserstoffatome sind weniger problematisch für das Peptidrückgrat eines Proteins (hier ist die Position und Ausrichtung eindeutig festgelegt). Sie sind jedoch sehr wohl relevant, um Protonierungszustände zu definieren und um Wasserstoffatome in Aminosäureseitenketten, deren Ausrichtungen nicht eindeutig festgelegt sind, zu beschreiben¹. Da CavKA die Interaktionserkennung nicht allein distanzbasiert zwischen schweren Atomen realisiert, sondern zusätzlich Winkelkriterien berücksichtigt, ist eine eindeutige Ausrichtung von Wasserstoffatomen erforderlich. Hierfür ist die MOE *Protonate-3D*-Methode in den Standardeinstellungen verwendet worden²²². *Protonate-3D* optimiert vollautomatisiert und ohne Einfluss des Anwenders die Ausrichtung und die Protonierungszustände von Ligand und Protein in einem Ligand-Rezeptor-Komplex. Abbildung 11.2 zeigt exemplarisch die Analyse von H-Brücken-Donor-Akzeptor-Interaktionen in einem Ligand-Rezeptor-Komplex.

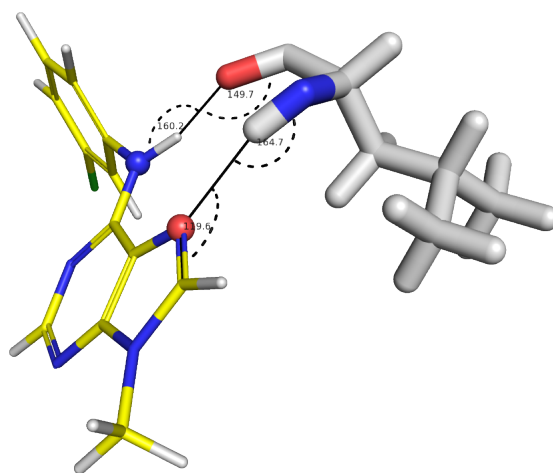


Abbildung 11.2: H-Brücken-Donor-Akzeptor-Interaktionen mit dazugehörigen Winkeln am Beispiel der *Cyclinabhängigen-Kinase 2* (PDB 1ckp). CavKA identifiziert eine H-Brücken-Interaktion in einer Distanz von 2.3-3.5Å (nicht gezeigt) und wenn die Winkelkriterien α (zwischen 110 und 180°; hier 160.2 und 164.7°) und β (zwischen 90 und 170°; hier 149.7 und 119.6°) erfüllt sind. Weitere Erläuterungen s. Text. Blaue Sphäre: interagierender H-Brücken-Donor des Liganden; rote Sphäre: interagierender H-Brücken Akzeptor des Liganden; gelbes Molekül: Ligand; grau: Aminosäure des Rezeptors).

Befinden sich innerhalb der vorgegebenen Distanz von 2.3-3.5Å kompatible Eigenschaftskombinationen (hier: H-Brücken-Donor und -Akzeptor), so werden nachfolgend die Winkel α und β geprüft, um sicherzustellen, dass Ligand- und Proteineigenschaften einander zugerichtet sind und eine Interaktion möglich ist.

Ein idealer Winkel α zwischen H-Brücken-Donoratom, H-Atom und H-Brücken-Akzeptoratom (Abk. Don-H...Acc) würde 180° betragen, da Donor-, Akzeptorfunktionalität und Wasserstoffatom sich linear zueinander anordnen. Ein optimaler Winkel β (zwischen einem H-Brücken-Akzeptor-Nachbaratom, einem H-Brücken-Akzeptor und einem H-Atom (Abk. X-Acc...H) würde 120° betragen, da sich ein Wasserstoffatom entlang der freien Elektronenpaare anordnet.

11.4 CavKA: Ionische Interaktionen / Akzeptor-Metall-Interaktionen

Ionische Interaktionen werden rein distanzbasiert zwischen positiv und negativ geladenen Atomen und unabhängig von H-Brücken-Interaktionen registriert. Selbiges gilt für Anionen-Metall- und Akzeptor-Metall-Interaktionen.

11.5 CavKA: Aromatische Interaktionen

Bei aromatischen Interaktionen wird in CavKA zwischen Kationen- π -, sog. 'Face-to-Face'- $\pi-\pi$ - (die Ringe sind parallel zueinander angeordnet) und sog. 'Edge-to-Face'- $\pi-\pi$ -Interaktionen (die Ringe sind senkrecht zueinander angeordnet) unterschieden. Die Implementierung wird nachfolgend erläutert.

11.5.1 Kationen- π -Interaktionen

In Abbildung 11.1 sind die geometrischen Voraussetzungen für Kationen- π -Interaktionen dargestellt. Hierbei wird ausgehend vom aromatischen Ringzentrum des Liganden ober- und unterhalb der Ringebene in 2.1Å⁶² II ein Punkt projiziert. Eine Kationen- π -Interaktion wird erkannt, wenn sich aromatisches Zentrum und kationische Eigenschaft in einer Distanz zwischen 2.5 und 4.5Å voneinander befinden. Der Winkel α zwischen aromatischem Ligand-Zentrum, aromatisch projiziertem Punkt und dem kationischen Zentrum muss hierbei zwischen 125° und 180° betragen.

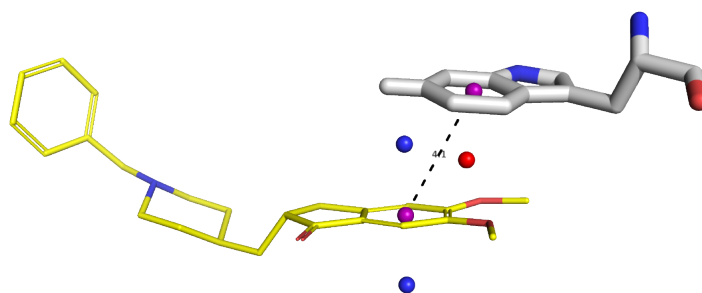
^{II}MOE verwendet ebenfalls eine Distanz von 2.1Å für projizierte Punkte in Pharmakophor-Modellen ausgehend vom aromatischen Zentrum⁶².

Bei projizierten Punkten (ober-, bzw. unterhalb der Ringebene) wird stets derjenige Punkt in der Berechnung berücksichtigt, der dem zu untersuchenden Zentrum (in diesem Falle die kationische Ligandeigenschaft) am nächsten ist.

11.5.2 '*Face-to-Face*'- $\pi - \pi$ -Interaktion

Bei einer '*Face-to-Face*'- $\pi - \pi$ -Interaktion zweier annähernd parallel angeordneter Ringe richten sich diese leicht versetzt zueinander aus. Das Wasserstoffatom des 'Donor'-Rings richtet sich hierbei idealerweise zentral über dem 'Akzeptor'-Ring aus (vgl. Abb. 11.3)¹²². Als Distanzkriterium für eine zustande kommende '*Face-to-Face*'- $\pi - \pi$ -Interaktion ist ein Toleranzbereich von 3.0-6.0Å zwischen den aromatischen Ringzentren gewählt worden (vgl. Abb. 11.3). Für aromatische Ligand- und Proteineigenschaften wurden in 2.1Å Entfernung Punkte ober- und unterhalb der Ringebene projiziert, wobei stets jener Punkt für die Winkelberechnung verwendet wurde, der dem zu untersuchenden aromatischen Zentrum am nächsten ist. Eine '*Face-to-Face*'- $\pi - \pi$ -Interaktion wird erkannt, wenn (vgl. Abb. 11.3):

- die Distanz zwischen den aromatischen Zentren (violett) zwischen 3.0 und 6.0Å beträgt (vgl. Abb. 11.3a; hier: 4.1Å),
- der Winkel α zwischen aromatischem Ligand-Ringzentrum (violett), projiziertem Ligand-Punkt (blau) und aromatischem Protein-Ringzentrum (violett) zwischen 120° und 180° (vgl. Abb. 11.3b; hier: $\alpha=126.5^\circ$) beträgt,
- der Winkel β zwischen aromatischem Proteinzentrum (violett), projiziertem Protein-Punkt (rot) und aromatischem Ligandzentrum (violett) weniger als 15° von α abweicht ($\delta < 15^\circ$) (vgl. Abb. 11.3c; hier: $\beta=131.3^\circ$; $\delta=4.8^\circ$).



(a) Distanzkriterium (hier: 4.1Å)

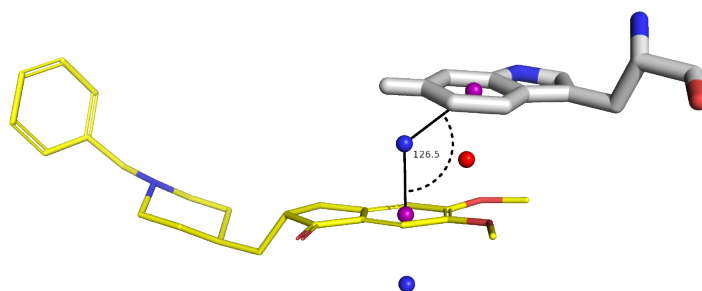
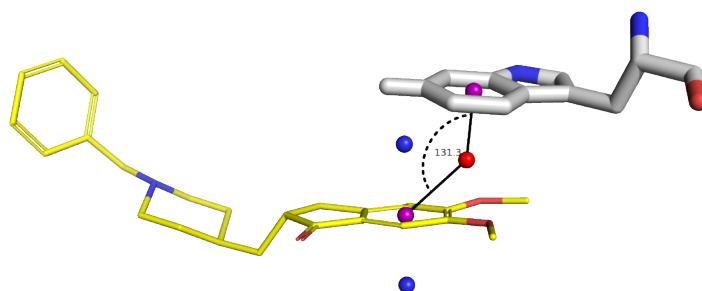
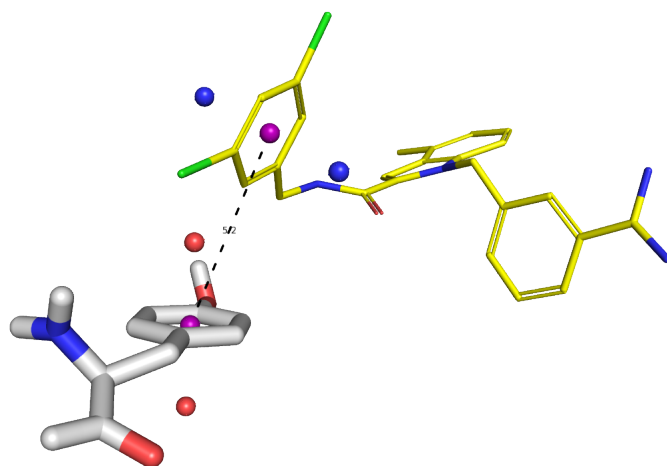
(b) Winkelkriterium α (hier: $\alpha=126.5^\circ$)(c) Winkelkriterium β (hier: $\beta=131.3^\circ$)

Abbildung 11.3: 'Face-to-Face'- $\pi - \pi$ -Interaktion (annähernd) parallel angeordneter aromatischer Ringe sowie die analysierten geometrischen Parameter. Weitere Erläuterungen s. Text. Gelb: Ligand (Donepezil); grau: Protein (Acetylcholinesterase; PDB 1eve), violett: aromatische Ringzentren, blau: projizierte Ligand-Punkte, rot: projizierter Protein-Punkt.

11.5.3 'Edge-to-Face'- $\pi - \pi$ -Interaktionen

Bei einer 'Edge-to-Face'- $\pi - \pi$ -Interaktion zweier (annähernd) senkrecht zueinander angeordneter aromatischer Ringe zeigt das interagierende Wasserstoffatom idealerweise in das Zentrum des 'Akzeptor'-Ringes (vgl. Abb. 11.4)¹²². Als Distanzkriterium für eine zustande kommende 'Edge-to-Face'- $\pi - \pi$ -Interaktion ist ein Toleranzbereich von 3.0-6.0Å zwischen den aromatischen Ringzentren gewählt worden (vgl. Abb. 11.4a). Für aromatische Ligand- und Proteineigenschaften wurden in 2.1Å Entfernung Punkte ober- und unterhalb der jeweiligen Ringebenen projiziert, wobei stets jener Punkt für Winkelberechnungen berücksichtigt wurde, der dem zu untersuchenden aromatischen Zentrum am nächsten ist. Eine 'Edge-to-Face'- $\pi - \pi$ -Interaktion wird erkannt, wenn (vgl. Abb. 11.4):

- die Distanz zwischen den aromatischen Zentren (violett) zwischen 3.0- 6.0Å beträgt (vgl. Abb. 11.4a; hier: 5.2Å),
- der Winkel β zwischen aromatischem Ligand-Ring-Zentrum (violett), projiziertem Protein-Punkt (rot) und aromatischem Aminosäure-Ring-Zentrum (violett) zwischen 140°-180° (vgl. Abb. 11.4b; hier $\beta=148.8^\circ$) beträgt,
- der Winkel γ zwischen projiziertem Ligand-Punkt (blau), aromatischem Ligand-Ring-Zentrum (violett) und aromatischem Aminosäure-Ring-Zentrum zwischen 80°-100° (vgl. Abb. 11.4c; hier: $\gamma=85.3^\circ$) beträgt.



(a) Distanzkriterium (hier: 5.2Å)

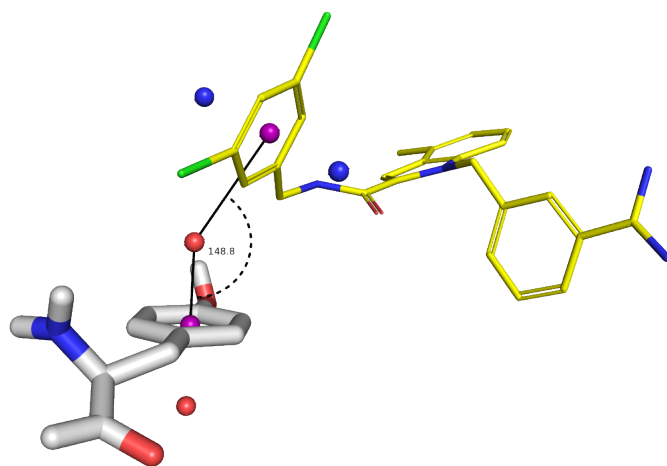
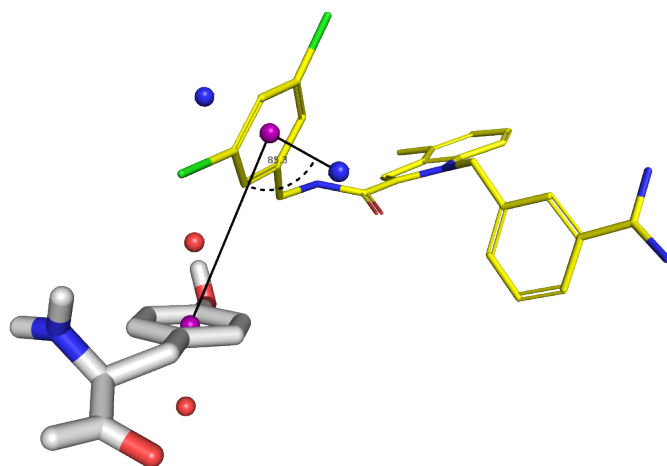
(b) Winkelkriterium β (hier: $\beta=148.8$)(c) Winkelkriterium γ (hier: $\gamma=85.3^\circ$)

Abbildung 11.4: 'Edge-to-Face'- $\pi - \pi$ -Interaktion zweier (annähernd) senkrecht angeordneter aromatischer Ringe sowie die analysierten geometrischen Parameter. Weitere Erläuterungen s. Text. Gelb: Ligand; grau: Protein (Faktor Xa; PDB 1lpz), violett: aromatische Ringzentren, blau: projizierte Ligand-Punkte, rot: projizierte Protein-Punkte.

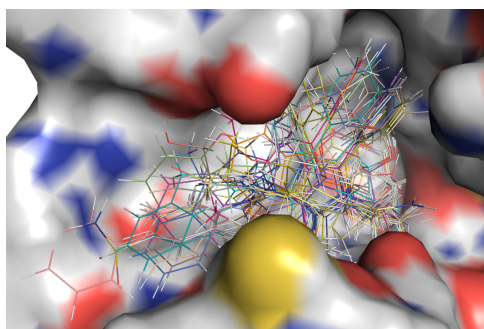
Kapitel 12

CavKA: Implementierung der energetischen Bindetaschen-Analyse

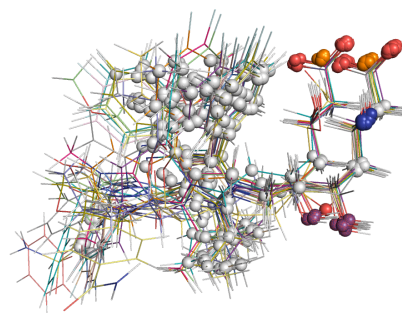
Die in Kapitel 3 beschriebenen theoretischen Hintergründe der Ligand-Rezeptor-Interaktion zeigen, wie viele unterschiedliche Faktoren die Affinität eines Liganden zu seiner Zielstruktur maßgeblich beeinflussen. All jene Kriterien einzig und allein mit geometrischen Regeln erklären zu wollen, scheint schier unmöglich. In diesem Kapitel wird erläutert, wie Methoden zur *Hotspot*-Detektion als eine wertvolle Ergänzung zu den geometrischen Methoden verwendet werden können, um strukturbasierte Pharmakophor-Modelle zu erstellen.

12.1 Auswahl der Hotspot-Methode

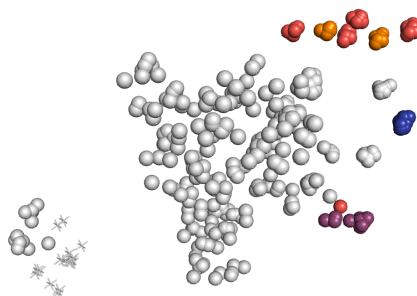
In Vorversuchen wurden gitter- und fragmentbasierte Methoden auf ihre Fähigkeit hin untersucht, *Hotspots* in Proteinbindetaschen in Abwesenheit von Ligandinformation detektieren zu können. Um Areale, in denen bevorzugt Interaktionen auftreten, zu identifizieren, wurden mehrere bereits superpositionierte Ligand-Rezeptor-Komplexe derselben Zielstruktur aus der Screening-PDB²³¹ (Abk. sc-PDB) verwendet (vgl. Abb. 12.1a), bei denen nach den unter in Kapitel 11 beschriebenen Methoden Interaktionen zwischen Ligand und Bindetasche berechnet wurden (vgl. Abb. 12.1b). Wenn viele superpositionierte Komplexe verwendet wurden, konnten in der Regel Areale in der Bindetasche identifiziert werden, in denen es gehäuft zu Interaktionen kam (Sphären in Abb. 12.1c), bzw. in denen es zu keiner Interaktion kam (Kreuze in Abb. 12.1c).



(a) 19 superpositionierte *HMG-CoA-Reduktase-Inhibitoren* in der Bindetasche von PDB 1hwi.



(b) Mit der Bindetasche interagierende Ligand-eigenschaften als Sphäre dargestellt.



(c) An einer Ligand-Rezeptor-Interaktion beteiligte Ligandeigenschaften als Sphäre, nicht-interagierende Eigenschaften als Kreuz dargestellt.

Abbildung 12.1: Identifikation von *Hotspots* am Beispiel von 19 superpositionierten *3-Hydroxy-3-Methylglutaryl-Coenzym-A-Reduktase-(HMG-CoA-Reduktase)-Inhibitoren* aus Ligand-Rezeptor-Komplexen aus der scPDB²³¹. *Blau*: *H-Brücken-Donor*-; *rot*: *H-Brücken-Akzeptor*-; *blau&rot*: *H-Brücken-Donor- und H-Brücken-Akzeptor*- ; *grau*: *hydrophobe*; *orange*: *anionische Eigenschaften*.

So wurde GRID²³² als gitterbasierter und MOE-MFSS⁶² als fragmentbasierter Ansatz miteinander verglichen. Es wurde untersucht, inwieweit beide Methoden in der Lage sind, die zuvor berechneten Interaktionsareale in Abwesenheit von Ligandinformation und in Abhängigkeit des Konturierungsgrades so *sensitiv* und *spezifisch* wie möglich wiederzufinden.

Abbildung 12.2 zeigt schematisch die Vorgehensweise. Werden die Felder, bzw. Fragmente zu *schwach* konturiert (vgl. Abb. 12.2a), werden neben den eigentlich relevanten Arealen (*richtig-positiv*; Abk. TP) auch fälschlicherweise viele andere Areale als wichtig vorhergesagt, in denen keine Interaktionen zu beobachten sind (*falsch-positiv*; Abk. FP). In Abbildung 12.2b ist ein optimaler Konturierungsgrad dargestellt: Areale, in denen Interaktionen zu beobachten sind, werden korrekt vorhergesagt (TP); Areale, in denen keine Interaktionen zu beobachten sind, werden nicht vorhergesagt (*richtig-negativ*; Abk. TN). Eine zu starke Konturierung der Felder (vgl. Abb. 12.2c) führt dazu, dass Areale, in denen es zu einer Interaktion

kommt, nicht mehr vorhergesagt werden (*falsch-negativ*; Abk. FN). Bei dem Vergleich von GRID und MFSS wurden die Konturierungsgrade der einzelnen Sonden, bzw. Fragmente auf Sensitivität und Spezifität optimiert (vgl. Gl. 14.1 u. 14.2 in Kap. 16).

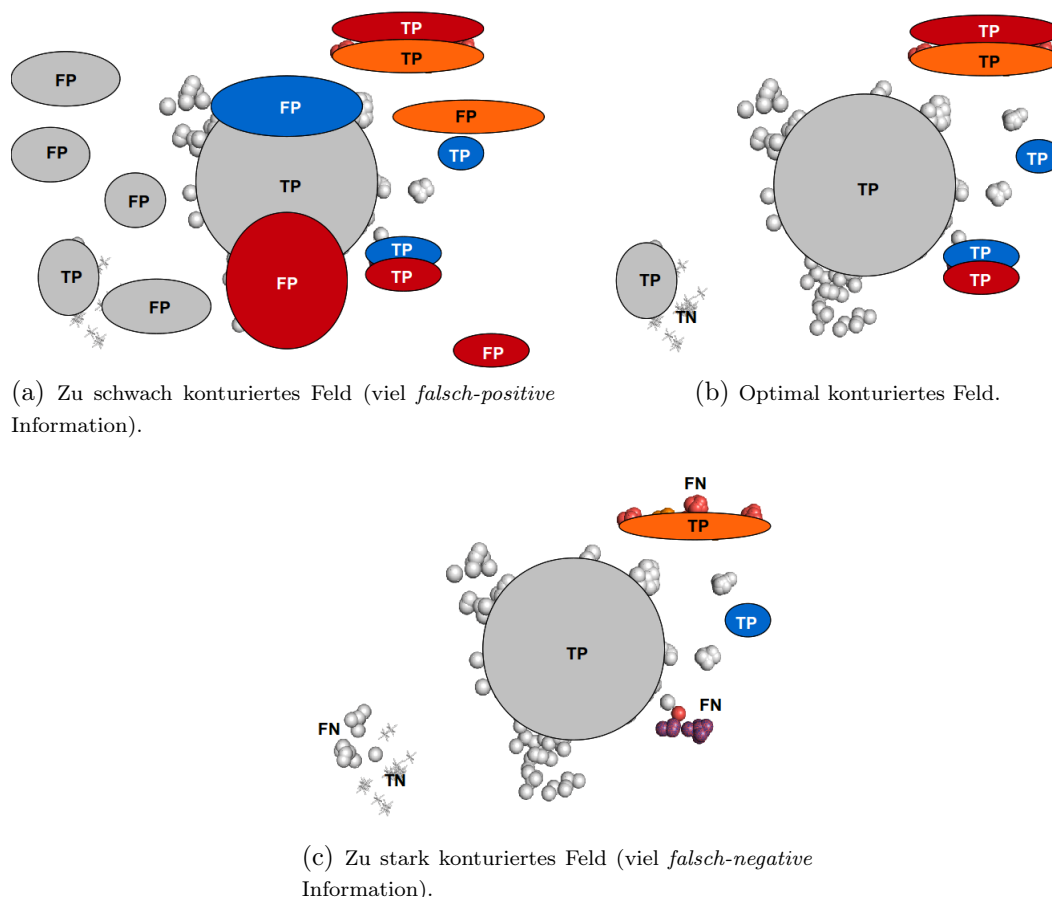


Abbildung 12.2: Schematische Darstellung der Optimierung des Konturierungsgrades. *Sphären*: interagierende Eigenschaften; *Kreuze*: nicht-interagierende Eigenschaften; *rot*: H-Brücken-Akzeptor-; *blau*: H-Brücken-Donor-; *orange*: anionische-; *grau*: hydrophobe Eigenschaften. *TP*: richtig-positiv; *FP*: falsch-positiv; *TN*: richtig-negativ; *FN*: falsch-negativ. Weitere Erläuterungen s. Text.

In Übereinstimmung mit der Arbeitsgruppe von KARPLUS, die in der Vergangenheit MCSS mit GRID verglich, erwiesen sich beide Methoden als annähernd gleichwertig²³³. Ein großer Vorteil fragmentbasierter Ansätze im Vergleich zu den gitterbasierten Methoden ist, dass sie die Ausrichtung eines Fragments berücksichtigen, also wie sich ein potentieller Ligand in der Bindetasche ausrichten könnte²³³. Diese Eigenschaft erweist sich vor allem im *de-novo Design* als sehr nützlich, da es oft nicht trivial ist, die *Hotspots* eines GRID-Feldes in entsprechende Ligand-Fragmente zu übersetzen.

Dennoch ist aus folgenden Gründen auf GRID als Methode zur *Hotspot*-Detektion in CavKA zurückgegriffen worden:

- Gitterbasierte Ansätze sind äußerst recheneffizient. Die Berechnung mehrerer MIFs einer Bindetasche bei einer Gitterauflösung von 0.5Å für mehrere Sonden erfordert Sekunden, wohingegen eine MFSS Rechnung mit 5000 Kopien^I pro Fragment Tage benötigt.
- Fragmentbasierte Ansätze tendieren dazu, sehr viele lokale Minima zu detektieren¹⁴⁸, deren visuelle Inspektion häufig schwer fällt. GRID-MIFs sind leichter zu interpretieren, da sie je nach Konturierungsgrad eine mehr oder minder große Fläche darstellen.

Im Folgenden wird beschrieben, wie MIFs idealerweise berechnet und vorbehandelt werden sollten, um *Hotspots* zu indizieren. Die in CavKA definierten Schwellenwerte zur MIF-Konturierung wurden ebenfalls in Mr. Brightsite verwendet.

12.2 GRID-Sonden

In GRID kann eine Vielzahl unterschiedlicher Sonden für nahezu jede pharmakophore Eigenschaft verwendet werden. So befinden sich in der aktuellen Version von GRID 77 unterschiedliche Sonden. Grob wird hierbei zwischen *einzelnen Sonden* (engl. *single-probes*) und *multiplen Sonden* (engl. *multiple-probes*) unterschieden. Bei *einzelnen Sonden* werden mehrere Atome implizit in einer Sonde berücksichtigt (z.B. der Wasser-Sonde, bei der Wasserstoffatome und Sauerstoffatom in *einer* Sonde berücksichtigt werden). Bei *multiplen Sonden* werden die einzelnen Atome explizit berücksichtigt.

In Übereinstimmung mit FOX und HAAKSMA zeigte sich in Vorversuchen, dass unterschiedliche Sonden, die ähnliche pharmakophore Eigenschaften besitzen, auch zu sehr ähnlichen MIFs führen¹⁵⁵, z.B. die O1- (aliphatische Hydroxylfunktion) und O- (phenolische Hydroxylfunktion) Sonden, was vermutlich auf eine ähnliche Sonden-Parametrisierung zurückzuführen ist. Um Redundanzen zu vermeiden, wurde in CavKA auf häufig verwendete Standard-Sonden zurückgegriffen (s. Tab. 12.1).

^IIn den MOE Standardeinstellungen werden 100 Kopien pro Fragment verwendet; in der Originalpublikation¹⁴⁵ wurden zwischen 1000 und 5000 Kopien pro Fragment verwendet.

E_{Grad} beschreibt die in CavKA verwendeten Konturierungsgrade der einzelnen Sonden, mit denen in Vorversuchen (vgl. Abschn. 12.1) die besten Ergebnisse erzielt worden sind. Abbildung 12.3 zeigt schematisch, wie MIFs für die CavKA-Methode vorbehandelt werden. Die Prozedur gliedert sich in drei Schritte und wird in den folgenden Abschnitten näher erläutert.

1. MIF-Berechnung mit einer Toleranz von $\pm 5\text{\AA}$ um den Liganden mit einer Gitterauflösung von 0.5\AA
2. Anpassung des MIFs an das Ligand-Areal mit einer Toleranz von 1.5\AA
3. Normierung und Konturierung der MIFs

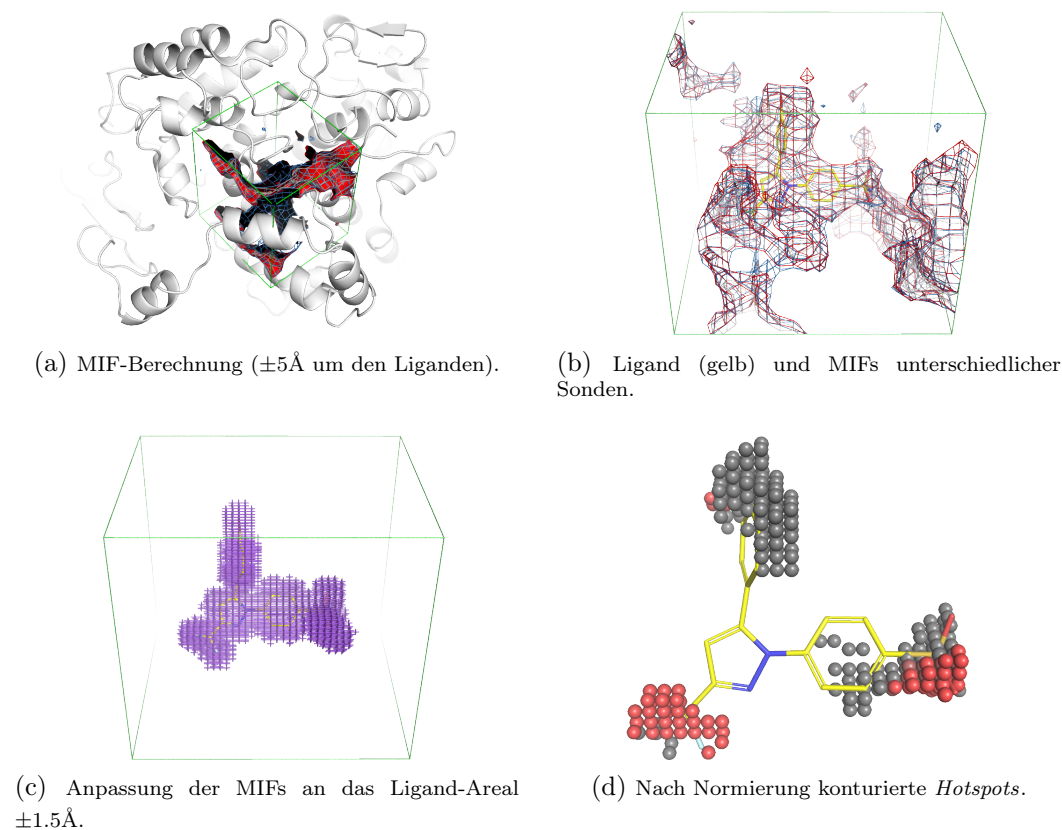


Abbildung 12.3: Prozedur der MIF-Prozessierung in CavKA am Beispiel der Cyclooxygenase 2 (PDB 1cx2). Rot: *N1-MIF*; blau: *O-MIF*; grau: *C3-MIF*.

Tabelle 12.1: Die für CavKA verwendeten GRID-Sonden und die verwendeten Konturierungsgrade.

Sonde	E_{Grad}	chemische Funktionalität	pharmakophore Eigenschaft
DRY	0.5	hydrophobe-Sonde	hydrophob
C3	0.7	Methyl-Gruppe	hydrophob/ sterisch
N1	0.7	amidischer Stickstoff	H-Brücken-Donor
O	0.6	sp^2 -Carbonyl-Sauerstoffatom	H-Brücken-Akzeptor
N1+	0.7	quart. sp^3 -Stickstoff	kationisch
O-	0.7	sp^2 -Phenolat-Sauerstoff	anionisch
H2O	0.9	Wasser	H-Brücken-Donor/ H-Brücken-Akzeptor

Abkürzungen: *Acc*: H-Brücken-Akzeptor; *Don*: H-Brücken-Donor; *Kat*: kationisch; *Ani*: anionisch; *Hyd*: hydrophob; E_{Grad} : Konturierungsgrad (Erläuterung s. Abschn. 12.6).

Die einzelnen Schritte der MIF-Prozessierung werden nachfolgend erläutert. Hierbei verwendete Parameter (Toleranzen für die MIF-Berechnung, Bindetaschendefinition, Skalierungs- und Konturierungsfaktoren) sind nur als Richtwerte zu verstehen, die individuell vom Anwender angepasst werden können.

12.3 MIF-Berechnung

Um ein MIF berechnen zu können, müssen zuvor die minimalen (x_{Bot} , y_{Bot} , z_{Bot}) und die maximalen (x_{Top} , y_{Top} , z_{Top}) Dimensionen des Gitters definiert werden, sodass sich ein Würfel, bzw. Quader (s. Abb. 12.3b) ergibt. In Vorversuchen erwies sich eine Gitterauflösung von 0.5\AA als zweckmäßig, um relevante Interaktionen ausreichend detailliert aufzulösen. Um sicherzustellen, alle relevanten Bereiche der Bindetasche adäquat abzubilden, sind die Gitterdimensionen relativ großzügig gewählt worden, sodass das Gitter den Liganden entlang jeder Achse um 5\AA überragt.

12.4 MIF-Adaption an das Ligand-Areal

GRID selbst bietet keine Möglichkeit, aus dem zuvor definierten Gitter die zur Bindetasche zugehörigen Areale herauszufiltern. Hierzu kann einer der zahlreichen publizierten Algorithmen zur Bindetaschendetektion verwendet werden. Für eine Übersicht über häufig verwendete Methoden siehe HENRICH et al.¹⁶⁸ und DREHER²²³. Die energetische Komponente CavKAs hat zum Ziel, die über den geometrischen Ansatz ermittelten interagierenden Ligandeneigenschaften zu priorisieren.

Somit werden bei CavKA nicht die Felder der ganzen Bindetasche betrachtet, sondern nur Gitterpunkte, die den Liganden mit einer Toleranz von 1.5\AA umgeben (s. Abb. 12.3c) unter der Voraussetzung, dass sie negative Interaktionswerte aufweisen. Bei der Implementierung von Mr. Brightsite zeigte sich, dass MIFs häufig viel *falsch-positive* Information beinhalten (s. Abschn. 6.4.1). Werden zu große Flächen, die gegebenenfalls viel falsch-positive Information enthalten, skaliert, kann es bei der sich anschließenden Konturierung (s. Abschn. 12.5) zu Verzerrungen der eigentlichen *Hotspots* kommen.

12.5 MIF-Skalierung

Die mit unterschiedlichen Sonden erhaltenen Interaktionswerte variieren zum Teil sehr stark. Analog zu den in Tabelle 3.1 genannten Werten werden mit unpolaren Sonden (Dry, C3) eher mäßige Interaktionswerte und mit ionischen (N1^+ , O^-) Sonden tendenziell eher starke Interaktionsenergien erhalten. Darüber hinaus besteht eine starke Abhängigkeit der Interaktionsenergien von der jeweils zu untersuchenden Zielstruktur, weshalb ein MIF-Abgleich über absolute Interaktionswerte wenig sinnvoll erscheint. Um eine Vergleichbarkeit der verwendeten Sonden *untereinander* sowie über die *unterschiedlichen* Zielstrukturen gewährleisten zu können, ist eine Skalierung der Interaktionsenergien erforderlich, deren Notwendigkeit bereits von KASTENHOLZ et al. im Rahmen ihrer GRID-/ CPCA-Methode ausführlich diskutiert worden ist²³⁴.

In der vorliegenden Arbeit wurde das von DREHER vorgeschlagene Konzept der MIF-Skalierung übernommen²²³. Abbildung 12.4 zeigt ein normiertes Histogramm über die relativen Häufigkeiten der Interaktionsenergien einer O-Sonde, an dem exemplarisch das Vorgehen wie folgt näher erläutert wird: Für jede Sonde wurde das 98% Perzentil ermittelt. Die Anzahl der 2% Punkte mit der höchsten Interaktionsenergie (die Spanne P-S) wurde mit 1 gewichtet. Alle anderen Punkte mit einem niedrigeren Interaktionspotential wurden linear abfallend von <1 bis 0 gewichtet.

Dieses Vorgehen ist zweckmäßig, um Verzerrungen in der sich anschließenden Konturierungsprozedur vorzubeugen: Innerhalb der MIFs kommt es immer wieder zu Artefakten ohne Informationsgehalt (isoliert auftretende Gitterpunkte mit einem extrem negativen Interaktionspotential, wohingegen die Nachbarpunkte ein sehr viel positiveres Potential besitzen).

Würde der Punkt mit der höchsten Interaktionsenergie P als Referenz für die Skalierung verwendet werden (und es sich hierbei um ein Artefakt handeln), würden als Folge dessen alle weiteren Punkte abgewertet werden. Die Gefahr, Artefakte oder falsch-positive Punkte zu erhalten, steigt mit der Ausdehnung des Gitters, weshalb mit 1.5\AA bewusst ein sehr kleiner Toleranzbereich gewählt worden ist.

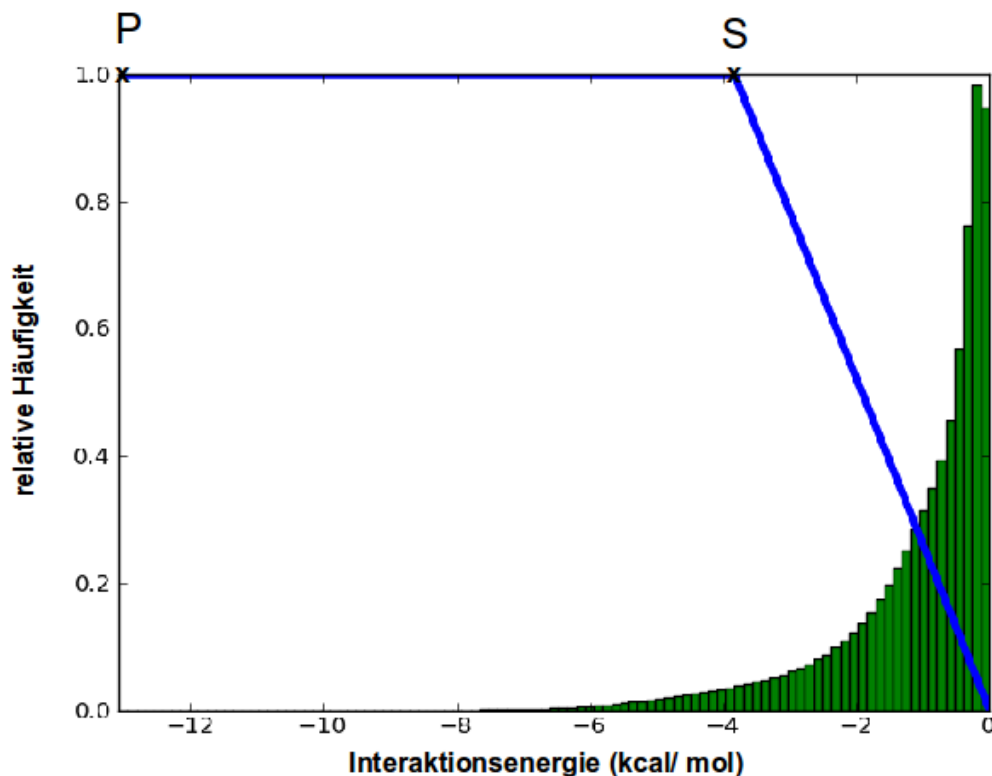


Abbildung 12.4: MIF-Skalierung am Beispiel einer O-Sonde. Der Übersicht halber wurde die relative Häufigkeit normiert. Abbildung verändert nach DREHER²²³. Erläuterungen s. Text.

12.6 MIF-Konturierung

Nachdem die Interaktionswerte der Gitterpunkte skaliert worden sind, werden die in Tabelle 12.1 genannten Schwellenwerte verwendet, um MIFs zu konturieren und so *Hotspots* zu indizieren (vgl. Abb. 12.3d). Ein Schwellenwert von 0.7 bedeutet, dass 70% der Punkte mit dem niedrigsten Interaktionspotential nicht betrachtet werden. Die 30% verbleibenden Punkte mit dem höchsten Interaktionspotential werden als *Hotspots* in Betracht gezogen.

Die automatisierte Konturierung der Felder stellt einen besonders kritischen Faktor in der CavKA-Methode dar. Das Definieren eines Schwellenwertes ist stets eine Gratwanderung zwischen der Generierung überflüssiger Information (Rauschen; *falsch-positive* Information; vgl. Abb. 12.2a), bzw. dem Nicht-Erkennen essentieller Interaktionen (*falsch-negative* Information; vgl. Abb. 12.2c).

Wann immer MIFs zur *Hotspot*-Detektion angewandt worden sind, um Pharmakophor-Modelle abzuleiten, sind die verwendeten Schwellenwerte in der Literatur empirisch und vor allem manuell festgelegt worden^{156,165,167}. Dies mag für einzelne Zielstrukturen sinnvoll erscheinen - nicht aber für eine Methode, die generell anwendbar sein soll.

Die ermittelten Schwellenwerte sind erste Erfahrungswerte sich dieser Problematik anzunähern. Sie sind keinesfalls universell gültig. Gewählte Schwellenwerte stellen erste Richtwerte dar, die individuell angepasst werden können, bzw. angepasst werden müssen.

Kapitel 13

Die CavKA-Software

Nachfolgend wird ein Überblick über die Möglichkeiten der CavKA-Software gegeben, Pharmakophor-Modelle nach unterschiedlichen methodischen Ansätzen abzuleiten. Die Beschreibung der durchgeführten Validierungsexperimente zur Leistungsbeurteilung folgt in Kapitel 16, die Validierungsexperimente der ligandenzentrischen Ansätze werden in Kapitel 17 näher beschrieben.

13.1 Manuelle Zusammenführung geometrischer und energetischer Kriterien

Das Konzept CavKAs besteht aus der Kombination eines geometrischen und energetischen Ansatzes, die sich gegenseitig ergänzen. Auf Basis der nach den geometrischen Regeln ermittelten interagierenden Eigenschaften wird ein Pharmakophor-Modell abgeleitet und gespeichert. Der Anwender kann optional konturierte MIFs laden und anhand der Kongruenz von interagierenden Eigenschaften und MIF ein Pharmakophor-Modell gegebenenfalls *manuell* modifizieren und anpassen.

Aktuell bietet CavKA die Möglichkeit, Modelle in das MOE⁶²-Format (.ph4; Harte-Sphären) sowie das Pharao²⁰³-Format (.phar; Gauß-Sphären) zu exportieren. Erstellte Modelle können anschließend in MOE oder Pharao geladen und für Datenbanksuchen verwendet werden. Abbildung 13.1 zeigt eine Übersicht der in CavKA verfügbaren Funktionalitäten. Das Fließdiagramm in Abbildung 25.1 im Anhang zeigt die zugrundeliegenden Module der CavKA Software, bzw. deren Einsatzzweck. Eine detaillierte Beschreibung der Module, eine Auflistung verwendeter frei verfügbarer Software sowie deren Integration in die CavKA-Software findet sich im Anhang in den Abschnitten 25.1 und 25.3. Die genauen, für die Validierung verwendeten Parameter werden in Abschnitt 16.1 erläutert.

geometrischer Ansatz

Identifikation von
Ligand-Rezeptor-Interaktionen
nach geometrischen Kriterien

(vgl. Kap. 11)

energetischer Ansatz

→ MIF-Berechnung
→ MIF-Anpassung an das
Ligand-Areal
→ MIF-Normierung
→ MIF-Konturierung

(vgl. Kap. 12)

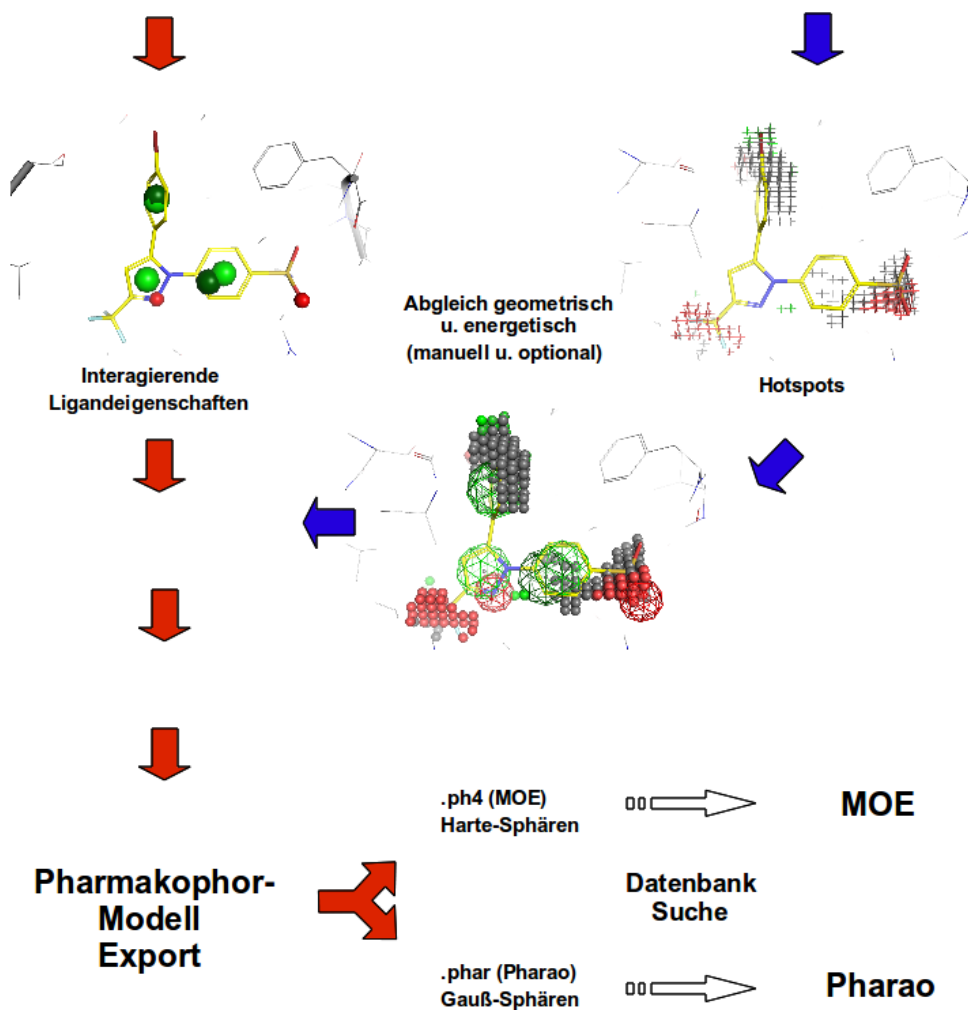


Abbildung 13.1: Schematische Darstellung der CavKA-Software. Auf Basis von Ligand-Rezeptor-Komplexen werden Ligand-Rezeptor-Interaktionen nach geometrischen Kriterien identifiziert, die in ein Pharmakophor-Modell exportiert werden können. Aktuell wird das MOE .ph4 und das Pharao .phar Format unterstützt. Optional können prozessierte MIFs (*Hotspots*) für die Modell-Optimierung verwendet werden. *Rote Pfeile: geometrischer Ansatz; blaue Pfeile: energetischer Ansatz; weiße Pfeile: extern zu verwendende Software; PDB: 1cx2 (Cyclooxygenase-2).*

13.2 MIF-basierte E-Pharmakophore

SALAM et al. präsentierten mit ihrem E-Pharmakophor-Ansatz (vgl. Abschn. 6.2.1) eine Methode, die Ligandeigenschaften in einem Ligand-Rezeptor-Komplex über eine Docking-Bewertungsfunktion (*Glide-XP*^{163,235}) priorisiert^{164,236}. Mit der Methode erstellten die Autoren Pharmakophor-Modelle, die keine nach geometrischen Kriterien identifizierten Ligand-Rezeptor-Interaktionen für die Hypothese berücksichtigten. Stattdessen wurde der Ligand-Rezeptor-Komplex nur energetisch über die Docking-Scoring-Funktion bewertet. Anstatt die priorisierten Ligandeigenschaften manuell in ein Pharmakophor zu überführen, wurde von den Autoren ein fester Regelsatz definiert, der die bewerteten Ligandeigenschaften direkt in ein Pharmakophor-Modell exportiert.

Dieser Regelsatz wurde adaptiert und GRID anstatt der *Glide-XP*-Scoring-Funktion verwendet, um MIF-basierte E-Pharmakophore zu erstellen. Aufgrund der mangelnden Fähigkeit, *Hotspots* punktgenau vorhersagen zu können, wurden statt kompletter MIFs, nach Kapitel 12 prozessierte MIFs verwendet. Würden die MIF-Punkte eines nicht konturierten Feldes genutzt werden, die der Ligandeigenschaft am nächsten sind, bestünde die Gefahr, irrelevante MIF-Interaktionswerte zu betrachten, wohingegen relevante Bereiche sich ggf. in der unmittelbaren Nähe befänden. So wurde den Koordinaten jeder einzelnen Ligandeigenschaft eine Gruppe von MIF-Punkten mit korrespondierenden pharmakophoren Eigenschaften (z.B. hydrophobe Ligandeigenschaft - GRID C3 Sonde) zugeordnet, wobei einer der Punkte aus dem MIF-Cluster maximal 1 Å von der betrachteten Ligandeigenschaft entfernt sein durfte. Erfolgte eine Zuordnung einer Ligandeigenschaft zu einer Gruppe von MIF-Punkten, wurde der arithmetische Mittelwert der Interaktionsenergien gebildet und der Ligandeigenschaft zugeordnet. Anhand der gemittelten Interaktionsenergien ließen sich die Eigenschaften priorisieren und wurden in eine Pharmakophor-Hypothese übersetzt. Eine Eigenschaft wurde als 'essentiell' betrachtet, wenn der Mittelwert der *skalierten* Interaktionsenergien des zugeordneten MIF-Clusters mindestens 0.90 kcal/mol beträgt. Erfolgt die Zuordnung von drei bis vier MIF-Clustern zu den jeweiligen Ligandeigenschaften (Pharmakophor-Hypothese mit drei, bzw. vier Sphären), müssen mindestens drei Eigenschaften des Modells von einem Datenbank-Molekül erfüllt werden. Bei fünf oder mehr Sphären müssen mindesten vier Eigenschaften des Modells erfüllt werden, wobei die maximale Anzahl der Sphären auf sechs begrenzt ist. Tabelle 13.1 gibt einen Überblick über den angepassten Regelsatz sowie den Regelsatz der Originalpublikation. Die genauen, für die Validierung verwendeten Parameter werden in Abschnitt 16.2 erläutert.

Tabelle 13.1: Übersicht über den modifizierten, bzw. originalen E-Pharmakophor-Regelsatz.

Regel	MIF-basiertes E-Pharmakophor	E-Pharmakophor nach SALAM ¹⁶⁴
energetische Definition	nach Kap. 12 prozessierte MIFs	Verwendung absoluter Energien
essentielle Sphären	skalierte Interaktionsenergie ≥ 0.90 kcal/mol	Glide XP Score ≤ -1.0 kcal/mol
a) partiell erfüllte Hypothesen	min. 3 (3-4 Sphären)	min. 3 (3-4 Sphären)
b) partiell erfüllte Hypothesen	min. 4 (≥ 5 Sphären)	min. 4 (≥ 5 Sphären)
maximale Sphären	max. 6	max. 6

13.3 CavKA^{HYBRID}

Auch wenn CavKA dem Anwender eine wertvolle Hilfe ist, bleibt die Erstellung und die Validierung eines Pharmakophors ein zeitraubender Prozess. Im Gegensatz zu Pharmakophor-Modellen, die eine Harte-Sphären-Repräsentation von Eigenschaften nutzen, besitzt die Repräsentation durch Gauß-Sphären den Vorteil, dass deutlich weniger Einfluss des Anwenders erforderlich ist.

CavKA^{HYBRID} kombiniert strukturbasierte und ligandzentrische Ansätze: Über die CavKA-Methode werden nach geometrischen Kriterien (vgl. Kap. 11) wichtige Schlüsseleigenschaften identifiziert und in ein Pharao-Pharmakophor-Modell exportiert. Die Pharmakophore werden, ohne verbotene Sphären zu berücksichtigen, für eine Datenbanksuche genutzt. Hierbei wird Pharao als Screening-Methode verwendet. Anstatt einer Hitliste werden die auf das Pharmakophor superpositionierten Datenbank-Moleküle in eine .sdf Datei exportiert. Diese wird erneut eingelesen und die Atom-Koordinaten der Datenbank-Moleküle gegen die Koordinaten der verbotenen Sphären abgeglichen. Befindet sich ein Atom des superpositionierten Moleküls näher als eine gerade noch tolerierte Distanz Tol_{Dist} vom Zentrum einer verbotenen Sphäre entfernt, wird für jedes dieser Atome der Score des superpositionierten Moleküls um den Strafterm $Tan_{Penalty}$ abgewertet. In CavKA^{HYBRID} sollen die Vorzüge einer Gauß'schen Superpositionierung (wenig Intervention des Anwenders erforderlich) mit denen eines Harten-Sphären-Modells (spezifische Modelle; in der Regel niedrigere *falsch-positiv* Raten) vereint werden. Der Bestrafungsterm sowie die tolerierte Annäherung an die verbotenen Sphären können vom Anwender festgelegt werden. Die genauen, für die Validierung verwendeten Parameter werden unter Abschnitt 16.3 erläutert. Energetische Kriterien werden derzeit bei Hybrid-Modellen nicht berücksichtigt.

Kapitel 14

CavKA: Validierung

14.1 Modellvalidierung und Leistungskennzahlen

Wann immer ein Modell (beispielsweise ein Pharmakophor-Modell, ein Ligand für einen ligandzentrischen Ansatz oder eine Bindetasche für eine Dockinganalyse) für ein Virtuelles Screening genutzt wird, ist dessen Qualität (also die Fähigkeit, Moleküle mit biologischer Aktivität anzureichern) von größtem Interesse. Bevor jedoch prospektiv nach neuen aktiven Substanzen gesucht wird, werden Modelle üblicherweise retrospektiv validiert. Dazu werden Moleküle, bei denen biologische Aktivität bereits bekannt ist (sogenannte Aktive; engl. *actives*) in einer weitaus größeren Datenbank mit biologisch inaktiven Molekülen (sogenannten Köder-Molekülen; engl. *decoys*) verteilt (engl. *sampling*). Ein gutes Modell ist in der Lage, die aktiven Moleküle von den inaktiven wieder zu unterscheiden.

Die Validierung von Pharmakophor-Modellen (insbesondere bei den Harten-Sphären-Modellen) ist ein *iterativer Prozess*. Es sind deutlich mehr Parameter zu berücksichtigen als bei einem ligandzentrischen Ansatz, z.B.:

- die Auswahl welche pharmakophoren Eigenschaften verwendet werden,
- der Sphärendurchmesser,
- Akzeptanzkriterien für partiell erfüllte Hypothesen (engl. *partial match*),
- Eigenschaften, die zwingend erfüllt sein müssen (engl. *essentials*),
- Definition verbotener Sphären (engl. *excluded volumes*).

Bei derartig vielen zu berücksichtigenden Parametern werden in den seltensten Fällen mit dem erstbesten Modell ideale Ergebnisse zu erzielen sein. Eine Pharmakophor-Hypothese ist nur eine Hypothese - was wörtlich zu verstehen ist: Sie muss kontinuierlich getestet, verändert oder gegebenenfalls auch verworfen werden⁷⁶. Daher sind in dieser Arbeit stets *optimierte Pharmakophor-Modelle* verwendet worden. Diese repräsentieren in den meisten Fällen lokale Optima, die keinesfalls dem globalen Optimum entsprechen müssen. In anderen Worten: Bei einem auf Basis *eines* Ligand-Rezeptor-Komplexes erstellten Modell ist der Raum für eine Pharmakophor-Optimierung begrenzt. Es ist hingegen möglich mit einem Modell, dass auf Basis eines anderen Ligand-Rezeptor-Komplexes erstellt worden ist, bessere Ergebnisse zu erzielen.

Abbildung 14.1 zeigt schematisch den iterativen Prozess der Pharmakophor-Optimierung. In den folgenden Abschnitten wird die Modell-Validierung verschiedener Screening-Methoden beschrieben.

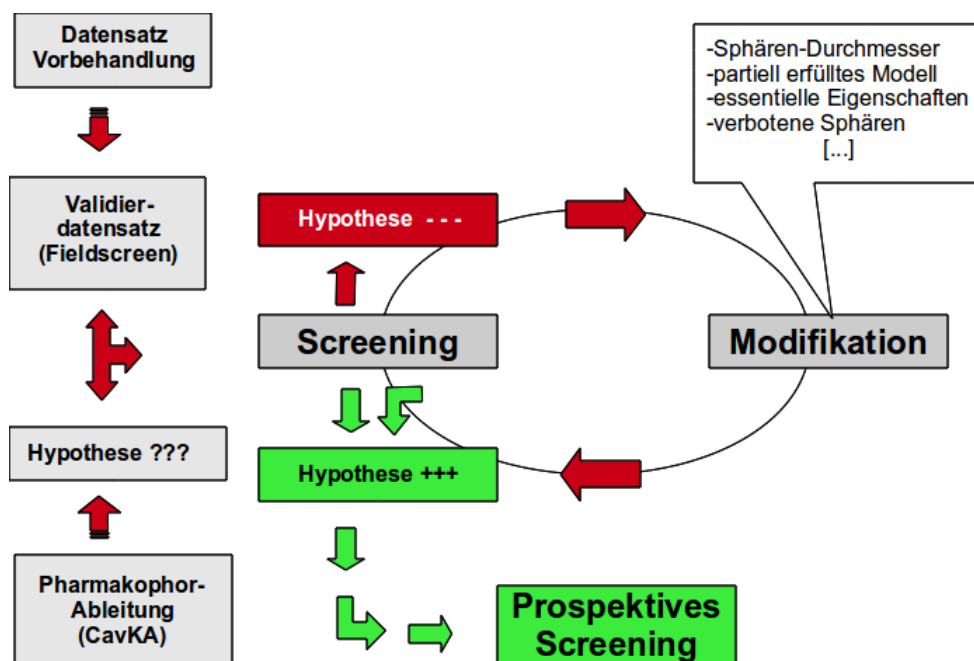


Abbildung 14.1: Schematische Darstellung der iterativen Pharmakophor-Optimierung. Über Methoden zur Pharmakophor-Ableitung werden Pharmakophor-Modelle erstellt, deren Fähigkeit aktive Moleküle anzureichern noch nicht bekannt ist. Die Leistungsfähigkeit eines Modells wird in einem retrospektiven Virtuellen Screening auf einem Validier-Datensatz analysiert. Modelle mit einem schlechten Anreicherungsverhalten werden modifiziert und weiter getestet. Modelle mit guter Leistungsfähigkeit werden häufig in prospektiven Screenings verwendet.

14.1.1 Validierung von Harte-Sphären-Modellen

Ein Hartes-Sphären-Pharmakophor-Modell klassifiziert Moleküle aus dem Validier-Datensatz in eine Fraktion, die das Modell *räumlich* und der *pharmakophoren Eigenschaft* nach erfüllt, bzw. nicht erfüllt. Die Entscheidungsgrenzen sind bei diesen Modellen sehr hart, da nur zwischen 'Pharmakophor erfüllt' (das Molekül wird in der Hitliste aufgeführt) und 'Pharmakophor nicht erfüllt' (das Molekül wird verworfen) unterschieden wird. Man spricht auch von 2-Klassen- oder binärer Klassifikation. Richtig klassifizierte Aktive werden als *richtig-positiv* (engl. *true positives*; Abk. TP), falsch klassifizierte Köder-Moleküle, die das Modell trotzdem erfüllen, als *falsch-positiv* (engl. *false positives*; Abk. FP) gewertet. Korrekterweise nicht gefundene Köder-Moleküle werden als *richtig-negativ* (engl. *true negatives*; Abk. TN), fälschlicherweise nicht gefundene Aktive als *falsch-negativ* (engl. *false negatives*; Abk. FN) gewertet. Abbildung 14.2 zeigt das schematische Vorgehen.

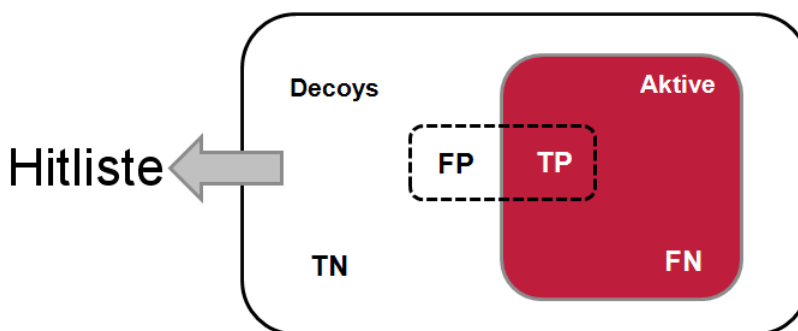


Abbildung 14.2: Schematische Darstellung einer Validierung. Eine Untermenge biologisch aktiver Moleküle (Aktive, rot) wird in eine deutlich größere Untermenge inaktiver Moleküle (Decoys, weiß) eingestreut. Eine Methode selektiert eine Untermenge (gestrichelte Linie) von n Molekülen aus einer Datenbank mit N Molekülen (Menge mit schwarz durchgezogener Linie) in einer Hitliste (verändert nach KIRCHMAIR et al.²³⁷; weitere Erläuterungen s. Text). *TP*: richtig-positiv; *FP*: falsch-positiv; *TN*: richtig-negativ; *FN*: falsch-negativ.

Aus diesen vier Kennzahlen lassen sich *Sensitivität* (als Maß für das Wiederfinden aktiver Moleküle) und *Spezifität* (als Maß für das Nicht-Finden der Köder-Moleküle) berechnen.

$$\text{Sensitivität} = \frac{TP}{TP + FN} \quad (14.1)$$

$$\text{Spezifität} = \frac{TN}{TN + FP} \quad (14.2)$$

Beide Leistungskennzahlen rangieren zwischen null und eins. Ist die Summe von *Sensitivität* und *Spezifität* größer als 1.0, ist davon auszugehen, dass eine bessere Anreicherung erzielt worden ist, als dies beim *zufälligen* Selektieren von Molekülen aus dem Validier-Datensatz zu erwarten gewesen wäre. Ein perfektes Modell wäre in der Lage, alle aktiven Moleküle (*Sensitivität* = 1.0) und kein Köder-Molekül zu finden (*Spezifität* = 1.0). Beide Kennzahlen können in der *Balanced Labeling Performance* (Abk. $Perf_{Bal}$) zusammengefasst werden, die den Mittelwert aus *Sensitivität* und *Spezifität* bildet.

$$Perf_{Bal} = \frac{Sensitivität + Spezifität}{2} \quad (14.3)$$

Bei der Interpretation der oben genannten Kennzahlen ist stets darauf zu achten, dass erheblich mehr Köder-Moleküle als Aktive vorhanden sind: eine Verbesserung der *Spezifität* um einen kleinen Betrag kann bedeuten, dass hunderte von Köder-Molekülen ausgeschlossen werden und sich die Hitliste dramatisch verkürzt. Bei allen Analysen, in denen *Sensitivität* und *Spezifität* zur Beurteilung ihrer Leistungsfähigkeit verwendet worden sind, wurden bevorzugt manuell optimierte Modelle zur finalen Bewertung herangezogen, die eine hohe *Spezifität* und somit eine relativ kurze Hitliste aufwiesen. Dieses Vorgehen soll dem Umstand Rechnung tragen, dass die Testkapazitäten (v.a. im akademischen Umfeld) im Allgemeinen stark begrenzt sind. Man würde also eine kurze Hitliste bevorzugen, in der man mit großer Sicherheit biologisch aktive Moleküle findet - auf Kosten einer höheren *falsch-negativ* Rate. Jene Überlegungen sind jedoch einzig und allein von den finanziellen Möglichkeiten sowie den verfügbaren Testkapazitäten bestimmt. *Sensitivität* und *Spezifität* sind gut als Gütemaße geeignet, wenn eine binäre Klassifizierung in aktiv / inaktiv (wie z.B. bei Harten-Sphären Pharmakophor-Modellen) vorgenommen wird.

Beim Betrachten von *Sensitivität* und *Spezifität* wird sehr schnell deutlich, dass beide Gütekriterien sich nicht unabhängig voneinander optimieren lassen: So ist es z.B. ein Leichtes, extrem spezifische Modelle zu erstellen, die von keinem der Köder-Moleküle, aber auch von keinem der aktiven Moleküle erfüllt werden (*Sensitivität* = 0.0; *Spezifität* = 1.0) oder extrem sensitive Modelle zu kreieren (*Sensitivität* = 1.0; *Spezifität* = 0.0), die so generisch sind, dass sie alle Aktiven, aber auch alle Köder-Moleküle anreichern.

14.1.2 Validierung von Gauß-Modellen

Im Gegensatz zu den oben genannten Methoden wird bei ligandenzentrischen Methoden oder Gauß-Modellen das Maß an Überlappung bezüglich der codierten Eigenschaft zwischen einem Referenz- und einem Datenbank-Molekül ermittelt. Alle besprochenen Programme bieten den *Tanimoto-Koeffizienten* (Abk. *TC*)^I als Gütemaß an, um die Ähnlichkeit zwischen den zu vergleichenden Molekülen zu bewerten:

$$TC = \frac{O_{A,B}}{O_{A,A} + O_{B,B} - O_{A,B}} \quad (14.4)$$

Wobei $O_{A,B}$ für die gemeinsame Überlappung der jeweils betrachteten Eigenschaft zwischen den zu vergleichenden Molekülen A und B steht und $O_{A,A}$, bzw. $O_{B,B}$ für die Überlappung mit sich selbst. Der *TC* kann Werte zwischen null und eins annehmen. Je nachdem, für welche der betrachteten Eigenschaften der *TC* berechnet worden ist, gelten unterschiedliche Annahmen die Ähnlichkeit betreffend, wobei generell gilt: Je größer der *TC*, desto größer ist die Ähnlichkeit.

Wenn eine Methode Moleküle über das Volumen abgleicht, ist ab einem $TC > 0.7$ molekulare Ähnlichkeit klar erkennbar¹⁸⁶. Wird mit Elektrostatik als Deskriptor gearbeitet, ist darauf zu achten, dass ein elektrostatisches Potential sowohl positive als auch negative Vorzeichen besitzen kann. In Folge dessen kann das Maß an Überlappung negative Werte annehmen. Bei der Verwendung von Elektrostatik kann ab einem *TC* von 0.2 von gewisser, ab einem *TC* von 0.4 von starker Ähnlichkeit¹⁸⁶ ausgegangen werden^{II}. Im Gegensatz zu der Hitliste eines Harten-Sphären-Modells wird eine kontinuierliche Hitliste erhalten, die anhand ihres *TC*s vom besten zum schlechtesten Wert geordnet wird.

^IÜblicherweise wird der *Tanimoto-Koeffizient* verwendet, um binäre Moleküldeskriptoren miteinander zu vergleichen. Ein Bit wird auf '1' gesetzt, sofern die betrachtete Eigenschaft in beiden zu vergleichenden Molekülen anzutreffen ist. Bei Abwesenheit der Eigenschaft wird das entsprechende Bit des Deskriptors auf '0' gesetzt. Bei Gauß-Modellen wird das Maß an Überlappung der Gauß-Volumina (dem Integral eines 3D-Gauß-Peaks) von einem Referenz- und einem Datenbank-Molekül abgeglichen.

^{II}In vielen Publikationen wird ein Tanimoto-Koeffizient von Molekülen ≥ 0.85 mit gleicher biologischer Aktivität gleichgesetzt. Diese Annahme entstammt Ergebnissen von Analysen, die mit UNITY Fingerprints durchgeführt und von vielen anderen Arbeitsgruppen aufgegriffen worden sind³. MARTIN et al. betonen in aktuellen Untersuchungen, dass es nicht möglich ist, anhand binärer Moleküldeskriptoren übergreifende und pauschale Tanimoto-Koeffizienten als Schwellenwerte zu definieren, die in der Lage sind, eine 'ähnliche' biologische Aktivität zu erklären²².

Eine solche geordnete Hitliste kann verwendet werden, um die Ergebnisse eines Virtuellen Screenings anhand einer sogenannten *Receiver Operator Characteristic* (Abk. *ROC*) zu interpretieren²³⁸. Die *ROC*-Kurve kann als ein ideales Qualitätsmaß angesehen werden, um die Fähigkeit einer Methode zu evaluieren, biologisch aktive Moleküle anzureichern²³⁹. Die *ROC*-Kurve vereint *Sensitivität* und (*1-Spezifität*), in einem einzigen Graphen (vgl. Abb. 14.3), ist frei von Wichtungsparemtern, und der Verlauf ist unabhängig vom Verhältnis Aktiver zu Köder-Molekülen im verwendeten Validier-Datensatz²³⁸. Vereinfacht zusammengefasst nehmen die Werte in *y*-Richtung beim Finden aktiver Moleküle zu, wohingegen beim Finden von Köder-Molekülen die *x*-Werte zunehmen. Im Idealfall würden sich nach einem Screening die aktiven Moleküle auf den vorderen Rängen der sortierten Hitliste befinden und die *ROC*-Kurve zu Beginn sehr steil ansteigen (blaue Linie). Das erklärte Ziel einer Validierung ist es, ein Modell zu finden, das aktive Moleküle in einem höheren Maße anreicht (rote Linie), als dies durch zufällige Selektion (gestrichelte Linie) der Fall wäre.

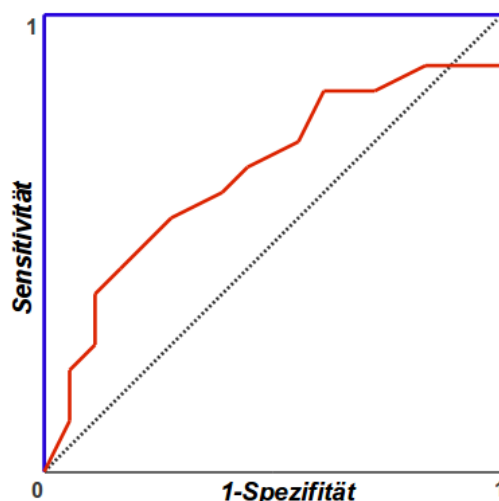


Abbildung 14.3: Schematische Darstellung einer ROC-Kurve. Blau - idealer Verlauf; rot - typischer Verlauf; schwarz gepunktet - zufälliger Verlauf (verändert nach TRIBALLEAU et al.²³⁸; weitere Erläuterungen s. Text).

Alternativ kann auch die Fläche unter der Kurve (engl. *area under the curve*) AUC_{ROC} betrachtet werden. Sowohl *ROC* als auch AUC_{ROC} wurden dahingehend kritisiert, dass eine Anreicherung aktiver Moleküle zu frühen Stadien der Datenbanksuche (engl. *early enrichment*) nur schlecht ersichtlich ist²³⁷. Eine frühe Anreicherung ist jedoch von entscheidender Relevanz, da üblicherweise Moleküle der vorderen Ränge in einem biochemischen Assay getestet werden.

Um dem Problem Rechnung zu tragen, entwickelte TRUCHON et al. mit der *Boltzman-Enhanced Discrimination of ROC* (Abk. *BEDROC*) ein Gütemaß, das einen variablen Wichtungsparemeter α nutzt, um die vorderen Ränge der Hitliste verstärkt zu gewichten und um somit eine Anreicherung in frühen Stadien besser darstellen zu können²⁴⁰. SHERIDAN et al. entwickelten das *Robust Initial Enhancement* (Abk. *RIE*), das die komplette Hitliste betrachtet und die steigenden Ränge exponentiell herunter gewichtet²⁴¹. Nachteilig ist bei *BEDROC* und *RIE*, dass weitere Parameter eingeführt werden, die definiert werden müssen und die Interpretation der Modelle erschweren.

In dieser Arbeit wurde die ROC-Kurve als Gütemaß für die Analyse kontinuierlicher Hitlisten gewählt, da keine benutzerdefinierten Parameter festgelegt werden müssen und die Kurve leicht interpretierbar ist. Um die Anreicherung, insbesondere auf den vorderen Rängen, besser erkennen zu können, ist bei allen in dieser Arbeit verwendeten ROC-Kurven eine logarithmierte Darstellung der x -Achse gewählt worden.

Ein weiteres, in dieser Arbeit verwendetes Qualitätsmaß sind sogenannte ROC-Anreicherungen (engl. *ROC enrichments*; Abk. ROC-EF). ROC-EF wird nachdrücklich von JAIN und NICHOLLS als Gütemaß zur Beurteilung eines Virtuellen Screenings empfohlen, da es sehr gut verwendet werden kann, um Anreicherungen auf den vorderen Rängen der Hitliste zu identifizieren²⁴². Darüber hinaus besteht keine Abhängigkeit des Gütemaßes vom Verhältnis Aktiver zu Köder-Molekülen, wie dies beim häufig verwendeten, sogenannten *Enrichment-Faktor* der Fall ist²⁴³.

$ROC - EF_{x\%}$ betrachtet die *Sensitivität* ($Sens_{FPx\%}$) bei einer bestimmten *falsch-positiv* Rate ($Spez_{FPx\%}$) bei $x\%$:

$$ROC - EF_{x\%} = \frac{Sens_{FPx\%}}{Spez_{FPx\%}} \quad (14.5)$$

Hierbei ist $Spez_{FPx\%}$ identisch mit $(1-Sens)$ auf der x -Achse der ROC-Kurve. In dieser Arbeit sind ROC-EF für 1%, 5% und 10% berechnet worden. Neben den hier genannten existiert eine Fülle alternativer Gütemaße^{244–246}. Einen Überblick über häufig verwendete Leistungskennzahlen sowie deren Stärken und Schwächen geben KIRCHMAIR et al.²³⁷.

An dieser Stelle ist noch einmal ausdrücklich darauf hinzuweisen, dass Modelle, die in einem Validierexperiment schlecht abschneiden, sehr wohl in der Lage sein können, prospektiv neue biologisch aktive Substanzen zu finden. Somit ist insbesondere das Datensatzdesign der verwendeten Validierdatensätze eine ganz besondere Herausforderung. Der Validier-Datensatz muss bezüglich der Eigenschaften, die in einem virtuellen Screening abgeprüft werden, *repräsentativ* für den später prospektiv zu durchsuchenden chemischen Raum sein. Weitere Anmerkungen zum Datensatzdesign finden sich in Abschnitt 14.3.

14.2 Generelle Probleme bei der Validierung strukturbasierter Pharmakophor-Modelle

In Abschnitt 14.1 sind die in dieser Arbeit verwendeten Gütemaße vorgestellt worden, die die Qualität von Pharmakophor-Modellen im Sinne der Anreicherung aktiver Moleküle in einem Virtuellen Screening beurteilen. In diesem Kapitel werden zusätzliche Kriterien besprochen, die bei der Modell-Erstellung (insbesondere bei Harten-Sphären-Modellen) sowie der Ergebnisinterpretation unbedingt zu berücksichtigen sind. Ferner wird auf Publikationen, bzw. Fallstudien verwiesen, in denen die jeweiligen Aspekte ausführlich diskutiert, bzw. Lösungsvorschläge gegeben werden. TRIBALLEAU et al. geben eine sehr gute Übersicht über Kernaspekte der Validierung von Pharmakophor-Modellen inklusive praktischer Anwendungsbeispiele⁹¹.

14.2.1 Überschießende Anzahl pharmakophorer Sphären

Methoden zur Ableitung von Pharmakophor-Modellen liefern Modellvorschläge, die potentiell wichtige Eigenschaften besonders berücksichtigen. In vielen Fällen wird jedoch eine sehr große Anzahl von Eigenschaften als relevant erachtet^{167,247} und (meist in nicht priorisierter Form) für eine Hypothese vorgeschlagen. Modelle mit sehr vielen Sphären bedürfen unbedingt einer Anpassung durch den Ausschluss von Sphären, bzw. der Akzeptanz von nur partiell erfüllten Modellen. Geschieht dies nicht, werden Modelle meist derart spezifisch, dass sie von keinem Datenbank-Molekül mehr erfüllt werden (*Sensitivität* = 0.0; *Spezifität* = 1.0). Die vorgeschlagene Anzahl der zu verwendenden Sphären wird kontrovers diskutiert: Während YANG eine Anzahl von maximal sieben Sphären vorschlägt²⁴⁷, empfehlen andere Autoren die Verwendung von üblicherweise drei, jedoch maximal vier Sphären²⁴⁸.

MIFs können in der CavKA Methode eine wertvolle Entscheidungshilfe sein, wenn nach geometrischen Kriterien sehr viele Eigenschaften als 'interagierend' erkannt werden und potentiell wichtig sind. In Abschnitt 18.1.3 wird gezeigt, wie bei einer überschießenden Anzahl pharmakophorer Eigenschaften MIFs dazu verwendet werden können, eine erste sinnvolle Vorauswahl von Eigenschaften zu treffen.

14.2.2 Ein Komplex - ein Bindungsmodus

Auch wenn 3D-Pharmakophor-Modellen häufig ein hohes Potential unterstellt wird, Moleküle mit unterschiedlichen Grundgerüsten in einem Screening zu finden⁴³, ist beim Erstellen strukturbasierter Pharmakophore auf Basis eines Ligand-Rezeptor-Komplexes zu beachten, dass ein einzelner Komplex in der Regel nur einen Bindungsmodus repräsentiert. Dieser Aspekt ist weniger relevant, sofern alle Liganden im gleichen Bindungsmodus binden. Existieren für eine Zielstruktur hingegen aktive Moleküle, die in unterschiedlichen Modi binden, so ist es sehr unwahrscheinlich, dass alle divers bindenden Liganden ein und denselben Pharmakophor erfüllen.

Das in Kapitel 15 vorgestellte *PLIF Driven Parallel Screening* (Abk. *PLIDriPaS*) stellt einen ersten Lösungsansatz dar, wie Ligand-Rezeptor-Komplexe identifiziert werden können bei denen Liganden in unterschiedlichen Modi binden. Jene Ligand-Rezeptor-Komplexe werden als Basis für die Erstellung von Pharmakophor-Modellen verwendet. Anstatt eine Datenbank mit einem Modell zu durchsuchen, wird ein Screening mit mehreren diversen Modellen durchgeführt.

14.2.3 Verbotene Sphären

Auch wenn in einigen Fällen Screening-Ergebnisse durch die Verwendung verbotener Sphären verbessert worden sind²⁴⁹, ist darauf zu achten, dass sie die Flexibilität des Proteins nicht berücksichtigen⁷⁶ und den Rezeptor fälschlicherweise als rigides Modell abbilden²⁵⁰. Viele unkritisch definierte verbotene Sphären können somit zu einer erhöhten *falsch-negativ* Rate führen⁹¹.

Mehrere Arbeitsgruppen versuchten daraufhin, verbotene Sphären nicht nur für eine einzige Rezeptor-Struktur zu definieren, sondern für ein ganzes über MD-Simulationen generiertes Rezeptor-Ensemble, um der Flexibilität des Rezeptors entsprechend Rechnung zu tragen^{156,251–253}. Alternativ kann der Sphärendurchmesser vergrößert werden. Auch wenn durch ein solches Vorgehen keine Rezeptorflexibilität im eigentlichen Sinne simuliert wird, trägt es dazu bei, weniger restriktive Modelle zu erstellen.

14.2.4 Methodenabhängigkeit der Suchergebnisse

Scheinbar identische Pharmakophor-Modelle (gleiche Anzahl sowie identische pharmakophore Ligandeigenschaften, gleiche Sphären-Durchmesser und gleiche Koordinaten), die mit unterschiedlichen Methoden generiert worden sind, können trotz ihrer vermeintlichen Ähnlichkeit unter Verwendung unterschiedlicher Programme zur Datenbanksuche zu stark unterschiedlich zusammengesetzten Hitlisten führen²¹⁵. Erklärungsansätze für dieses Phänomen begründen sich unter anderem in^{215,220}:

- unterschiedlichen Algorithmen der Superpositionierung von Datenbank-Molekül und Pharmakophor-Modell,
- unterschiedlichen Scoring-Funktionen, die die Superpositionierung bewerten,
- der Definition, wann eine verbotene Sphäre als verletzt gilt und zum Verwerfen des Datenbank-Moleküls führt.

Ferner spielt die Definition pharmakophorer Ligandeigenschaften eine entscheidende Rolle: Auch wenn die Eigenschaften im *Modell* identisch sind, definiert jede Methode die Eigenschaften der *Datenbank-Moleküle* unterschiedlich^{III}, was die Ergebnisse letztlich beeinflusst^{215,220}. Somit dürfen Ergebnisse nie allein auf ein Modell zurückgeführt werden, sondern müssen stets im engen Zusammenhang der Screening-Methode interpretiert werden²¹⁵.

14.3 Datensätze

Jedes Ergebnis, das im Rahmen einer retrospektiven Validierung erzielt worden ist, muss in Relation zum jeweils verwendeten Datensatz gesehen werden^{11,39,254–257}. Wenn Aktive und Köder-Moleküle nur unterschiedlich genug sind, wird es für nahezu jede Methode ein Leichtes sein, die aktiven Moleküle wieder von den Köder-Molekülen zu trennen.

VERDONK et al. stellten bei einer Evaluierung des Dockingprogramms GOLD fest, dass die erzielten Ergebnisse stark von der Zusammensetzung der Köder-Moleküle abhingen: Sie unterschieden sich anhand einfacher physikalisch-chemischer Eigenschaften, wie der molekularen Masse, der Anzahl der H-Brücken-Donoren und H-Brücken-Akzeptoren²⁵⁷.

^{III}Beispielsweise in der Definition hydrophober Eigenschaften; der Lokalisation einer anionischen Eigenschaft bei einer Carboxylat-Gruppe (am Sauerstoffatom mit der negativen Formalladung, bzw. zwischen den Sauerstoffatomen delokalisiert)^{214,220}.

Ein Datensatzdesign, das eine Separation Aktiver und Köder-Moleküle anhand einfacher Eigenschaften ermöglicht, führt häufig zu scheinbaren Anreicherungen (engl. *artificial enrichment*) und ist unter allen Umständen zu vermeiden. Zum einen sind keine anspruchsvollen Screening-Techniken notwendig, um gute Ergebnisse zu erzielen; zum anderen bilden etablierte Techniken, wie z.B. das Dockingprogramm GOLD, nur das mangelhafte Datensatzdesign ab²⁵⁷ - worauf nicht automatisch auf die Leistungsfähigkeit der Methode rückgeschlossen werden kann.

Nachfolgend werden drei Datensätze und deren zugrundeliegende Konzeption näher erläutert. Neben den hier genannten Validierdatensätzen wurden mehrfach große Anstrengungen unternommen, Datensätze zu entwickeln, die aufgrund ihrer Komposition scheinbaren Anreicherungen vorbeugen sollen^{257,258}.

14.3.1 Directory of Useful Decoys (DUD)

Directory of Useful Decoys (Abk. DUD) ist einer der derzeit größten frei verfügbaren Datensätze für strukturbasierte Screening-Methoden²⁵⁹. Der Datensatz umfasst 40 verschiedene, pharmazeutisch relevante Zielstrukturen, zu denen jeweils eine Röntgenkristallstruktur (bzw. Homologie-Modell) bekannt ist. Obwohl ursprünglich als Docking-Validier-Datensatz entwickelt, ist DUD^{182,185,191,243,260} oder eine Auswahl von DUD-Zielstrukturen (vgl. 14.3.3)^{133,184,261,262} als Validier-Datensatz für unterschiedliche ligandzentrische Ansätze verwendet worden.

Zu biologisch *aktiven Molekülen*, die aus einer Literaturrecherche stammen, wurden *Köder-Moleküle* aus der ZINC-Datenbank²⁶³, einer Datenbank kommerziell erhältlicher Substanzen, so gewählt, dass sie, bezogen auf einfache physikalisch-chemische Eigenschaften (molekulare Masse, logP-Wert, Anzahl der H-Brücken-Donoren, -Akzeptoren und rotierbarer Bindungen), den Aktiven sehr ähnlich sind. Zusätzlich bemühten sich die Autoren um eine ähnliche Verteilung häufig vorkommender funktioneller Gruppen wie Aminen, Amiden und Carbonsäuren. Die oben genannten Maßnahmen sollen einer scheinbaren Anreicherung entgegenwirken. Um auszuschließen, dass die ausgewählten Moleküle biologisch aktiv sind, war eine weitere Bedingung notwendig: Die Köder-Moleküle mussten sich von den Aktiven topologisch *unterscheiden*. Hierzu wurde ein Tanimoto-Koeffizient für einen 2D-Fingerprint (CAVCTVS-Typ2-Fingerprint²⁶⁴) definiert, der eine Mindest-Diversität ($TC < 0.9^{IV}$) zwischen Aktiven und Köder-Molekülen festlegte²⁵⁹.

^{IV}Ein mit CAVCTVS-Typ2-Fingerprint ermittelter TC von 0.9 soll einem mit Daylight-Fingerprints ermittelten TC von 0.7 entsprechen²⁵⁹.

Das DUD-Datensatzdesign kann in den folgenden zwei Punkten zusammengefasst werden:

- Aktive und Köder-Moleküle weisen eine *hohe Ähnlichkeit* bezüglich einfacher physikalisch-chemischer Eigenschaften auf (hohe 1D-Ähnlichkeit),
- Aktive und Köder-Moleküle weisen eine stark *unterschiedliche Topologie* auf (hohe 2D-Diversität).

GOOD und OPREA kritisierten, dass die aktiven Moleküle untereinander eine sehr hohe Ähnlichkeit aufweisen (engl. *analogue bias*)²⁶⁵. Eine starke Ähnlichkeit unter den aktiven Molekülen birgt ebenfalls die Gefahr scheinbarer Anreicherungen: Sobald ein Modell ein aktives Molekül detektiert, werden analoge Substanzen in den meisten Fällen ebenfalls gefunden. Um den Einfluss analoger Moleküle zu reduzieren, verwendeten die Autoren ein reduziertes Graphen-Clustering mit dem Ziel, Moleküle mit gleichen Grundgerüsten in Gruppen zusammenzufassen. GOOD und OPREA empfahlen eine Screening-Methode nach der Fähigkeit zu bewerten, möglichst viele unterschiedliche Grundgerüste zu detektieren. Sie betonten gleichzeitig, dass unterschiedliche Grundgerüst-Cluster dennoch eine beachtliche Ähnlichkeit aufweisen können²⁶⁵.

Ein weiterer Kritikpunkt betrifft die Bioaktivität der Köder-Moleküle. Diese sind als *inaktiv* definiert worden, sofern sie sich hinreichend von den Aktiven unterscheiden. Dennoch besteht die Möglichkeit, dass die vermeintlichen Köder-Moleküle biologisch aktiv auf der zu untersuchenden Zielstruktur sind (*falsch-negativ* Problematik)¹¹. Die *Inaktivität* der DUD-Köder-Moleküle ist nicht experimentell bestätigt worden. VENKATRAMAN et al.²⁶⁰ stellten ein extrem gutes Abschneiden von 2D-Methoden auf DUD fest. Diese Beobachtungen sollten keinesfalls überraschen: In jenen Analysen wird aufgrund der hohen 2D-Diversität zwischen Aktiven und Köder-Molekülen einzig und allein das DUD zugrundeliegende Datensatzdesign abgebildet. Dieses Phänomen wird im Folgenden als *DUD-2D-Bias* bezeichnet. DUD ist somit ungeeignet für Screening-Methoden, die Moleküle anhand ihrer 2D-Topologie vergleichen.

14.3.2 Maximum Unbiased Validation (MUV)

Um die genannten Mängel des DUD-Datensatzdesigns zu überwinden und einen für ligandbasierte Methoden geeigneten Validier-Datensatz zu entwickeln, stellten ROHRER und BAUMANN basierend auf PubChem-Daten²⁶⁶ mit *Maximum Unbiased Validation* (Abk. MUV) einen frei verfügbaren Datensatz zusammen¹¹. MUV umfasst 17 unterschiedliche, pharmazeutisch relevante Zielstrukturen. Für jede Zielstruktur wurden 30 diverse Moleküle verwendet, deren Aktivität in einem Bestätigungs-Assay festgestellt worden ist. Als Köder-Moleküle (15000 für jede Zielstruktur) wurden nur solche verwendet, die im Primär-Screening negativ getestet worden sind, was als Vorteil gegenüber DUD zu werten ist^V.

Dadurch ist es möglich, dass zu den diversen aktiven Molekülen, Köder-Moleküle bezüglich einfacher physikalisch-chemischer Eigenschaften so hinzugefügt werden, dass:

- die aktiven Moleküle möglichst gut in die Köder-Moleküle eingebettet sind,
- die Köder-Moleküle untereinander möglichst gleichmäßig verteilt sind.

Durch eine gleichmäßige Einbettung der aktiven Moleküle in die Köder-Moleküle soll einer scheinbaren Anreicherung vorgebeugt werden. Der Datensatz wurde aufgrund der ausgewählten Zielstrukturen (Reporter-Gen-Assays für fünf Zielstrukturen) sowie dem geringen Anreicherungsverhalten unterschiedlicher Screening-Techniken kritisiert²⁶⁷. Als Ursache für das geringe Anreicherungsverhalten diskutierten TIIKKAINEN et al. sprunghafte SARs, inkorrekt annotierte Bioaktivität der Köder-Moleküle sowie unterschiedliche Bindungsmodi der aktiven Moleküle an ihrer Zielstruktur²⁶⁷.

Da nur für einen kleinen Teil der MUV-Zielstrukturen kristallographische Daten frei verfügbar sind, wurde sowohl für die Validierung der CavKA-Methode als auch für den Vergleich der ligandzentrischen Methoden der im Folgenden beschriebene, sogenannte *Fieldscreen-Datensatz* verwendet.

^VIm Gegensatz zu DUD handelt es sich nicht nur um 'vermutete' inaktive Köder-Moleküle.

14.3.3 Fieldscreen-Datensatz

Aufgrund der mangelnden Diversität unter den aktiven Molekülen ist der DUD-Datensatz selbst nur begrenzt für Validierungsexperimente geeignet²⁶⁵. CHEESERIGHT et al. verwendeten deshalb für die Validierung ihrer ligandzentrischen Fieldscreen-Methode¹⁸⁴ nur jene DUD-Zielstrukturen, die mindestens 15 verschiedene der von GOOD und OPREA²⁶⁵ festgelegten Grundgerüst-Clustern enthielten. Die so erhaltene, 13 Zielstrukturen umfassende DUD-Auswahl wird im Folgenden als Fieldscreen-Datensatz bezeichnet und diene als Datengrundlage für die in dieser Arbeit durchgeführten Analysen.

Ein großer Vorteil des Datensatzes besteht darin, dass für elf Zielstrukturen kristallographische Daten hinterlegt sind. Aufgrund mangelnder Verfügbarkeit von Röntgenkristallstrukturen für die Zielstrukturen *Platelet Derived Growth Factor Receptor Kinase Beta* (Abk. PDGFRB) und *Vascular Endothelial Growth Factor Receptor 2* (Abk. VEGFR2) wurden Homologie-Modelle auf der Basis nahe verwandter Röntgenkristallstrukturen erstellt und auf der DUD-Homepage hinterlegt²⁶⁸. Als Grundlage für das PDGFRB-Homologie-Modell diene die *cKIT-Kinase* (PDB 1t46). Als Basis für das VEGFR2-Homologie-Modell diene der *Fibroblast Growth Factor Receptor 1* (PDB 1fgi)²⁵⁹. In der Fieldscreen-Publikation wurden diejenigen Ligand-Rezeptor-Komplexe zitiert, auf deren Basis die Homologie-Modelle erstellt worden sind (1t46 für PDGFRB; 1fgi für VEGFR2).

Da bei einem ligandzentrischen Ansatz ausschließlich Ligandinformation berücksichtigt wird, ist zu beachten: Bei dem co-kristallisierten Liganden aus PDB 1t46 handelt es sich um ein aktives Molekül für die Zielstruktur *cKIT-Kinase*. Bei den DUD-Aktiven für PDGFRB handelt es sich hingegen auch tatsächlich um PDGFRB-Inhibitoren. Der Referenz-Ligand und die Aktiven des Datensatzes können sich gegebenenfalls stark unterscheiden. Gleiches gilt in Analogie für VEGFR2.

Beim Docking spielt die oben genannte Problematik eine weniger große Rolle: Die DUD-Aktiven der Zielstrukturen PDGFRB und VEGFR2 können ohne weiteres in das entsprechende Homologie-Modell gedockt werden. In diesem Falle ist eine Ligand-Abhängigkeit sehr viel geringer als bei einem ligandzentrischen Ansatz. Nachfolgend werden die in der Fieldscreen-Publikation zitierten Ligand-Rezeptor-Komplexe als 'Original-Ligand-Rezeptor-Komplexe' bezeichnet.

Im Rahmen der CavKA-Validierung sollte analysiert werden, inwieweit relevante Ligand-Rezeptor-Interaktionen von der Methode erkannt werden, um Pharmakophor-Modelle mit einem guten Anreicherungsverhalten zu erstellen. Eine eindeutige Ligand-Rezeptor-Interaktionserkennung setzt korrekte und qualitativ hochwertige Ligand-Rezeptor-Komplexe voraus. Da die DUD-Autoren die Erstellung der

Homologie-Modelle nicht weiter beschreiben²⁵⁹ und die ursprünglichen Liganden aus PDB 1t46 und 1fgi letztlich keine Aktiven für diese Zielstrukturen darstellen, ist für die Ableitung der Pharmakophor-Modelle jeweils der Original Ligand-Rezeptor-Komplex verwendet worden. Die erzielten Anreicherungen für die Zielstrukturen PDGFRB und VEGFR2 sollten daher unter Vorbehalt interpretiert werden. Für die ligandzentrischen Ansätze ist für PDGFRB und VEGFR2 ebenfalls auf die co-kristallisierten Liganden aus dem Original Ligand-Rezeptor-Komplex zurückgegriffen worden: Diese sind identisch mit denen auf der DUD-Homepage hinterlegten Strukturen für PDGFRB²⁶⁹ und VEGFR2²⁷⁰.

Eine Übersicht über Zielstrukturen sowie die Anzahl aktiver Moleküle, Köder-Moleküle und Grundstruktur-Cluster des Fieldscreen-Datensatzes ist in Tabelle 14.1 zu finden.

Tabelle 14.1: Fieldscreen-Datensatz nach CHEESERIGHT et al.¹⁸⁴.

Ziel- struktur	Anzahl Cluster	Anzahl aktive Moleküle	Anzahl Köder- Moleküle	PDB Referenz ^{184,259}	Name Ligand
ACE	19	46	1718	1o86	Lisinopril
AChE	19	100	3654	1eve	Donepezil
CDK2	32	47	1747	1ckp	Purvalanol B
COX2	44	212	11218	1cx2	1-Phenylsulfonamide- 3-Trifluoromethyl- 5-Parabromophenylpyrazole
EGFR	40	365	14193	1m17	Erlortinib
FXa	19	64	1809	1f0r	RPR208815
HIV-RT	17	33	1382	1rt1	6-Benzyl-1-Ethoxymethyl-5- Isopropyl Uracil
INHA	23	56	2406	1p44	GENZ-10850
P38	20	137	5542	1kv2	BIRB 796
PDE5	22	26	1501	1xp0	Vardenafil
PDGFRB	22	124	5031	1t46	Imatinib
SRC	21	98	4906	2src	Phosphoaminophosphonic Acid-Adenylate Ester
VEGFR2	31	48	2389	1fgi	3-[(3-(2-Carboxyethyl)-4- Methylpyrrol-2-yl)Methylene]- 2-Indolinone

Abkürzungen (engl.): *ACE*: Angiotensin Converting Enzym; *AChE*: Acetylcholinesterase; *CDK2*: Cyclin Dependent Kinase 2; *COX2*: Cyclooxygenase 2; *EGFR*: Epidermal Growth Factor Receptor; *FXa*: Factor Xa; *HIV-RT*: HIV-Reverse Transkriptase; *INHA*: Enoyl-ACP-Reductase; *P38*: P38 Mitogen activated Protein; *PDE5*: Phosphodiesterase-5; *PDGFRB*: Platelet Derived Growth Factor Receptor Kinase Beta; *SRC*: Tyrosin Kinase SRC; *VEGFR2*: Vascular Endothelial Growth Factor Receptor 2.

Eine gute Methode im Virtuellen Screening sollte nicht nur der Anzahl nach möglichst viele Hits liefern, sondern idealerweise strukturell diverse Moleküle⁵³ (vgl. Abschn. 1.5) anreichern.

Um dies zu beurteilen, wurden alle in dieser Arbeit untersuchten Methoden nach dem *chemotype enrichment* bewertet²⁷¹, wobei die Leistungsfähigkeit einer Methode danach beurteilt wird, wie viele unterschiedliche Grundgerüste sie in der Lage ist zu detektieren. Hierbei ist auf die von GOOD und OPREA definierten Graphen-Cluster zurückgegriffen worden, um unterschiedliche Grundgerüste zu klassifizieren²⁶⁵.

Eine besondere Beachtung verdient hierbei die Zusammensetzung der einzelnen Grundstruktur-Cluster: In den Datensätzen existieren zahlreiche Cluster mit sehr vielen Mitgliedern, wohingegen andere nur ein einziges Mitglied besitzen. Würden die gefundenen Aktiven unabhängig von ihrer Cluster-Herkunft gleich gewichtet werden, so dominierten jene Cluster, die viele Mitglieder umfassen, die Screening-Ergebnisse sehr viel stärker als einzelne Cluster mit wenigen Mitgliedern. Deshalb wurden die Aktiven bei der Berechnung der in dieser Arbeit verwendeten Gütemaße *invers*^{VI} zu ihrer Clustergröße gewichtet^{13,271}. Die Cluster gewichtete *richtig-positiv* Rate $TP_{Cl-weighted}$ kann wie folgt berechnet werden:

$$TP_{Cl-weighted} = \sum_{i=1}^k \frac{1}{N_{actives}^i} \cdot N_{found}^i \quad (14.6)$$

Hierbei ist k die Anzahl aller Scaffold-Cluster, $N_{actives}^i$ die Gesamtanzahl aller aktiven Moleküle im jeweils betrachteten Scaffold-Cluster i und N_{found}^i die Anzahl der Moleküle, die durch eine Screening-Methode gefunden wurden. Hieraus kann die Cluster gewichtete *Sensitivität*¹³ $Sens_{Cl-weighted}$ wie folgt berechnet werden:

$$Sens_{Cl-weighted} = \frac{TP_{Cl-weighted}}{k} \quad (14.7)$$

Werden aktive Moleküle gefunden, die einem Cluster mit vielen Mitgliedern angehören, steigt die gewichtete Sensitivität langsam an, wohingegen das Finden

^{VI}Wenn eine Methode aus einem vier Mitglieder umfassenden Cluster ein Molekül findet, so erfolgt die Berechnung des zugrunde gelegten Gütemaßes nicht mit 1 TP , sondern mit $\frac{1}{4} = 0.25$ TP .

von Clustern, die nur ein Mitglied umfassen, zu einem sprunghaften Anstieg führt. MACKEY und MELVILLE merkten jedoch an, dass die statistische Robustheit mit der Anzahl an Mitgliedern pro Cluster zunimmt²⁷¹.

14.4 Daten-Vorbehandlungen

14.4.1 Zusammenstellung des Fieldscreen-Datensatzes

Als Datensatz für alle in dieser Arbeit durchgeführten Validierungsexperimente wurde der Fieldscreen Datensatz verwendet. Zunächst wurden die Datensätze aller Aktiven und Köder-Moleküle für die 13 Fieldscreen-Zielstrukturen des korrigierten DUD-Datensatzes in der zweiten Version heruntergeladen²⁶⁸. Für die Zusammenstellung des Datensatzes wurden nur die in den Zusatzinformationen der Fieldscreen-Publikation referenzierten ZINC IDs verwendet²⁷² VII. Die Clusterzugehörigkeit der aktiven Moleküle wurde aus den von GOOD erstellten .sdf Dateien *DUD Clusters*²⁷³ für die jeweiligen Zielstrukturen übernommen und auf die aktiven Moleküle im Fieldscreen-Datensatz übertragen^{VIII}. Falsche Bindungsordnungen wurden für die Moleküle mit den ZINC IDs 03833931, 03833932 und 03833937 (Aktive aus dem PDGFRB-Datensatz) manuell in MOE korrigiert.

Für die unterschiedlichen Validier-Studien (die Validierung strukturbasierter Pharmakophor-Modelle sowie der Vergleich der ligandenzentrischen Ansätze) erfolgte eine unterschiedliche Datensatz-Vorbehandlung der Aktiven und Köder-Moleküle, die nachfolgend erläutert wird. Konfigurationsdateien für verwendete Programme mit den entsprechenden Parametern sind im Anhang in Kapitel 27 aufgelistet.

^{VII}Es ist auf die Original-DUD-Dateien zurückgegriffen worden, da ZINC IDs für etliche Moleküle geändert wurden, bzw. unter vielen IDs keine Moleküle mehr hinterlegt sind. Die Verwendung der aktuellen ZINC-Datenbank ist somit wenig sinnvoll.

^{VIII}Von der Verwendung der *DUD Clusters* .sdf Dateien ist dringend abzuraten, da während der Prozessierung mit Pipeline Pilot eine große Anzahl der Strukturen falsch abgespeichert worden ist (fünfbändige Kohlenstoffatome; aromatische Zustände aufgehoben etc.). Dies hat zum einen völlig andere pharmakophore Eigenschaften zur Folge, zum anderen führt es in vielen Fällen zu Fehlern beim Einlesen der Dateien, woraufhin jene Moleküle in Hitlisten plötzlich nicht mehr aufgeführt sind.

14.4.2 Vorbehandlung der Aktiven und Köder-Moleküle

14.4.2.1 Für Screenings mit strukturbasierten Pharmakophor-Modellen

Die Vorbereitung der Datensätze für die strukturbasierten Pharmakophor-Modelle ist aufwändig. Bei der Erstellung solcher Modelle werden Ligand-Rezeptor-Komplexe als Grundlage verwendet. Tautomere Zustände eines Liganden können hierbei eine erhebliche Relevanz für Ligand-Rezeptor-Interaktionen besitzen^{274,275}. So wurden zunächst energetisch günstige Tautomere mit der QUACPAC-*Tautomers*-Methode²⁷⁶ erstellt^{IX}, von denen im Anschluss Konformeren-Ensembles mit OMEGA¹⁹⁹ generiert worden sind. OMEGA benutzt einen deterministischen, regelbasierten Ansatz für die Analyse des Konformationsraumes und hat sich bewährt, bioaktive Konformationen zu reproduzieren²⁷⁷.

Es wurden Konformere in unterschiedlichen Auflösungen berechnet:

1. maximal 100 Konformere pro Molekül bei einem Energiefenster von 25 kcal/mol (engl. *high-throughput-setting*; Abk. *HTS*; vgl. KIRCHMAIR et al.¹⁹⁸).
2. maximal 500 Konformere pro Molekül bei einem Energiefenster von 20 kcal/mol (engl. *high-quality-setting*; Abk. *HQS*; vgl. KIRCHMAIR et al.¹⁹⁸).

Mittels der QUACPAC-*pkaTyper*-Methode²⁷⁶ in den Standardeinstellungen wurden Protonierungszustände bei pH-Werten von 2-14 für die jeweiligen Konformere definiert: In einer Bindetasche können pH-Werte teils drastisch variieren, was die Protonierungszustände des Liganden beeinflusst und entscheidend für mögliche Ligand-Rezeptor-Interaktionen ist. Die auf diese Weise vorbehandelten Datensätze werden nachfolgend gemäß ihrer Konformeren-Auflösung als *struct_HTS* (niedrige Auflösung), bzw. *struct_HQS* (hohe Auflösung) abgekürzt.

14.4.2.2 Für Screenings mit ligandzentrischen Methoden

Bei ligandzentrischen Ansätzen wird die Zielstruktur nicht in der Datenbanksuche berücksichtigt. Vielmehr wird ein biologisch aktiver Referenz-Ligand mit den Datenbank-Molekülen abgeglichen. Die Datensatz-Vorbehandlung umfasste die Definition von Protonierungszuständen über die FILTER-*pkaNorm*²⁷⁸-Methode (in diesem Fall für einen pH-Wert von 7.4) sowie das Erstellen von Konformeren-Ensembles. Konformere wurden unter Verwendung von OMEGA¹⁹⁹ mit den o.g. HTS-Einstellungen generiert.

^{IX}Hierzu wurde die 'reasonable' Option in QUACPAC verwendet.

Für die mit *ShaEP* durchgeführten Analysen war zusätzlich die Definition von Partialladungen erforderlich. Dazu wurden die bereits vorbehandelten Datenbanken mittels BABEL²⁷⁹ in das .mol2 Dateiformat konvertiert und Partialladungen über die QUACPAC-*Molcharge*²⁷⁶-Methode berechnet (verwendet wurde das MMFF94 Kraftfeld in den Standardeinstellungen). Wie auch in den Analysen ligandzentrischer Ansätze anderer Autoren wurde auf die Definition tautomerer Zustände verzichtet^{133,184,185,191,192,260,261}. Der nach diesem Schema vorbehandelte Datensatz wird nachfolgend als *ligcent_HTS* bezeichnet.

14.5 Vorbehandlung der Screening Referenzen

Im folgenden Abschnitt wird beschrieben, wie die Screening-Referenzen (ein Ligand-Rezeptor-Komplex für die Erstellung strukturbasierter Pharmakophor-Modelle, bzw. ein co-kristallisierter Ligand für die ligandzentrischen Ansätze) vorbehandelt werden. Die Namen der Dateien, die im Rahmen der Vorbereitungs-Routinen größtenteils automatisiert erzeugt und abgespeichert worden sind, sind in Klammern aufgeführt.

14.5.1 Ligand-Rezeptor-Komplexe

Die in der Fieldscreen-Publikation¹⁸⁴ genannten Ligand-Rezeptor-Komplexe für die 13 Zielstrukturen wurden aus der Protein-Datenbank^{160,161} anhand ihres vierstelligen Identifikationscodes heruntergeladen. Von den Strukturen wurden jeweils nur Monomere eines Ligand-Rezeptor-Komplexes verwendet (Abk. *_mono_complex.pdb*) und mit der MOE *Protonate-3D*-Methode²²² (Abk. *_mono_complex_3D.pdb*) vorbehandelt. *Protonate-3D* definiert Protonierungszustände und Ausrichtung von Wasserstoffatomen ohne Einfluss des Anwenders sowohl für das Protein als auch für den Liganden. Der prozessierte Ligand wurde in einer separaten .mol Datei gespeichert (Abk. *refined_LIGAND_3D.mol*), Bindungsordnungen wurden ggf. manuell korrigiert. Anschließend wurden der Ligand, Gegenionen, Kristallisationsartefakte und Co-Faktoren aus dem Ligand-Rezeptor-Komplex entfernt. Für so vorbehandelte Proteine wurde eine Version mit Wassermolekülen (Abk. *refined_PROTEIN_3D_waters.pdb*) sowie eine Version ohne Wassermoleküle (Abk. *refined_PROTEIN_3D.pdb*) im .pdb Format gespeichert. Die auf diese Weise erhaltenen Ligand-Rezeptor-Komplexe bildeten die Grundlage für die komplexbasierten Pharmakophor-Ableitungen mit Ligandscout⁸⁸ und CavKA.

Für die MIF-Berechnungen mit GRID werden PDB-Dateien ohne Wassermoleküle sowie Wasserstoffatome benötigt. Letztere werden im Zuge der GRID-Berechnung durch das Programm selbst festgelegt. So wurden bereits prozessierte .pdb-Dateien mit BABEL²⁷⁹ und der '-h' Option in eine .pdb Datei ohne Wasserstoffatome konvertiert (Abk. *protein_NO_H.pdb*), die anschließend auf Konsistenz geprüft wurde (Entfernung von Nicht-Standard Aminosäuren; konsistente Nummerierung der Aminosäure-Reste; Abk. *protein_NO_H_clean.pdb*^X). Die oben genannten Prozessierungsroutinen wurden weitestgehend durch das Python-Programm *prepare_cavka.py* automatisiert.

^XDiese Vorbehandlung ist absolut essentiell für die GRID-Berechnungen. Kleinste Uneindeutigkeiten innerhalb der .pdb Datei führen zum Abbruch des Programms.

14.5.2 Referenz-Liganden

Die in der Fieldscreen-Publikation genannten Ligand-Rezeptor-Komplexe wurden aus der Protein-Datenbank^{160,161} anhand ihres vierstelligen Identifikationscodes heruntergeladen. Liganden wurden aus der Bindetasche extrahiert und visuell auf korrekte Bindungszustände geprüft. Falls nötig wurden diese in MOE⁶² manuell korrigiert und separat als MDL .mol-Datei gespeichert. Um einen fairen Vergleich der ligandzentrischen Methoden untereinander zu ermöglichen, wurde die FILTER Routine *pkaNORM*²⁷⁸ für die Definition von Protonierungszuständen des Referenz-Liganden bei einem physiologischen pH-Wert von 7.4 verwendet. Ein pH-Angleich von Referenz- und Datenbank-Molekül wird von ROCS¹⁹⁰ und EON²⁰⁰ automatisch vorgenommen. Für die Berechnung von elektrostatischen Feldern ist die Definition von Partialladungen erforderlich. Partialladungen werden bei EON ebenfalls automatisch durch das MMFF94-Kraftfeld in den Standardeinstellungen berechnet. Für ShaEP¹⁸⁵ wurde der Referenz-Ligand zunächst mittels BABEL²⁷⁹ in das .mol2 Format konvertiert und MMFF94-Partialladungen unter Verwendung der QUACPAC-*Molcharge*²⁷⁶-Methode berechnet.

14.6 Hitlistenprozessierung

Die primär von der jeweiligen Methode ausgegebenen Hitlisten enthalten noch verschiedene Konformere ein und desselben Moleküls und bedürfen weiterer Prozessierung (ansonsten würden beispielsweise für drei sich in der Hitliste befindende unterschiedliche Konformere, die auf dasselbe Molekül zurückzuführen sind, fälschlicherweise *drei* TP gezählt). Anhand ihrer ZINC-IDs wurden die Konformere gefiltert, sodass pro Molekül nur das am *besten* bewertete Konformer in der ROC-Kurve berücksichtigt wurde (engl. *best hit*).

Mit Pybel-Inchis²²⁸ wurde geprüft, ob sich der verwendete Referenz-Ligand unter den aktiven Liganden befindet. War dies der Fall, so wurden die entsprechenden Positionen in der Hitliste nicht berücksichtigt, um einer scheinbaren Anreicherung entgegenzuwirken.

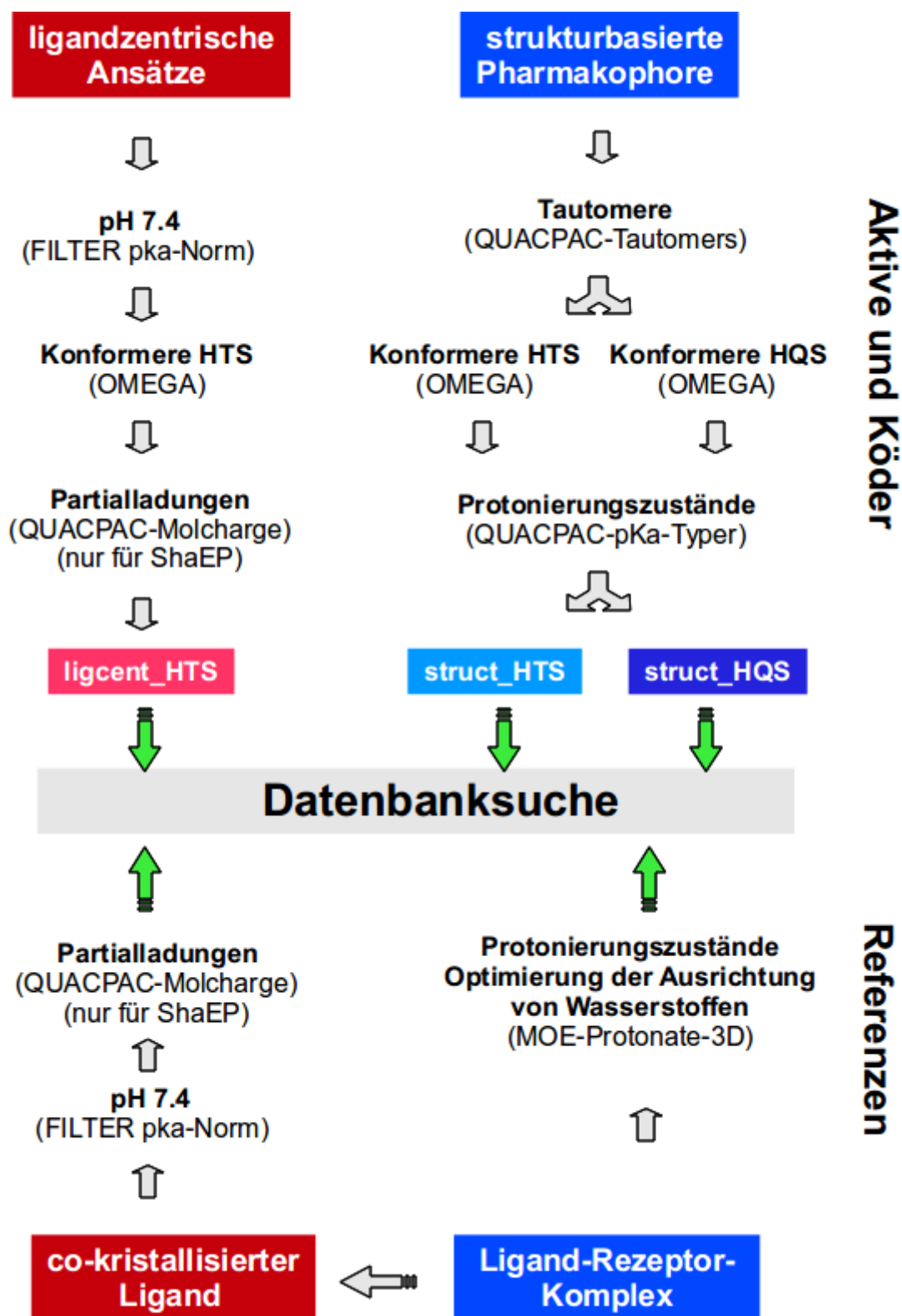


Abbildung 14.4: Übersicht über die einzelnen Schritte der Daten-Vorbehandlung sowie die verwendete Software für Datensätze und Referenzen. Weitere Erläuterungen s. Text.

Kapitel 15

PLIF Driven Parallel Screening (PLIDriPaS)

Es wurde bereits darauf hingewiesen, dass Pharmakophor-Modelle, die auf Basis eines Ligand-Rezeptor-Komplexes erstellt worden sind, nur repräsentativ für diesen Bindungsmodus sind. In anderen Worten: Es ist nahezu unmöglich mit einem solchen Modell aktive Liganden anzureichern, die einen völlig anderen Bindungsmodus zeigen.

Die hier beschriebene Methode *PLIF Driven Parallel Screening* (Abk. PLIDriPaS) kombiniert *Protein-Ligand-Interaktions-Fingerprints* (Abk. PLIF) als Werkzeug, um co-kristallisierte Liganden in unterschiedlichen Bindungsmodi in Ligand-Rezeptor-Komplexen derselben Zielstruktur zu identifizieren, mit der Grundidee des von STEINDL et al. vorgestellten *Parallel Screenings*, bei dem Datenbanksuchen parallel mit *unterschiedlichen* Pharmakophor-Modellen durchgeführt wurden²⁸⁰⁻²⁸².

15.1 Protein Ligand Interaktions Fingerprint (PLIF)

Structural Interaction Fingerprints (Abk. SIFt) war die erste publizierte Methode, die Interaktionen zwischen einem Liganden und den Aminosäuren der Bindetasche in einem Fingerprint codierte²⁸³. Ein sich hieraus ergebendes Anwendungsspektrum ist vielfältig und reicht vom Clustern von Docking-Posen, der Analyse verwandter Proteinfamilien auf Verwandtschaftsbeziehungen bis zur gezielten Suchen nach bestimmten Interaktionen in Docking-Posen, um z.B. eine erneute Bewertung durchzuführen (engl. *re-scoring*)²⁸⁴.

Analog zum SIFt wurden ähnliche Konzepte implementiert²⁸⁵⁻²⁸⁷, die sich im Wesentlichen nur in ihrer Auflösung unterscheiden, d.h., wie detailliert die Interaktionen zwischen Ligand und Bindetasche codiert werden. Erwähnenswert ist, dass in der frei zugänglichen *Screening-PDB*²³¹ einer annotierten PDB-Datenbank, bereits für jeden Ligand-Rezeptor-Komplex ein *Interaction Fingerprint* (Abk. IFP) hinterlegt ist²⁸⁶.

In der vorliegenden Arbeit wurde eine analoge Implementierung in MOE verwendet, der *Protein Ligand Interaction Fingerprint* (Abk. PLIF), um alternative Bindungsmodi co-kristallisierter Liganden in Ligand-Rezeptor-Komplexen zu identifizieren⁶².

Abbildung 15.1 zeigt die Strukturformeln von *Dihydrofolat* (Abk. DHF) und *Methotrexat* (Abk. MTX). Beide Moleküle binden an eine gemeinsame Zielstruktur, die *Dihydrofolat-Reduktase* (Abk. DHFR). Aufgrund der hohen strukturellen Ähnlichkeit der Liganden (*Tanimoto-Koeffizient*= 0.84, berechnet über *Openbabel Fingerprint 2*²¹⁷, dem Standard-Fingerprint in Openbabel) würde man keine unterschiedlichen Bindungsmodi erwarten.

Superpositioniert man jedoch die Ligand-Rezeptor-Komplexe PDB 1dhs (DHF) und PDB 1dls (MTX), stellt man das Gegenteil fest: DHF und MTX unterscheiden sich deutlich bezüglich der Anordnung pharmakophorer Eigenschaften. In Abbildung 15.1 sind jene Ligandeigenschaften farblich hervorgehoben, die unterschiedliche Interaktionen mit der Bindetasche ausbilden. Aufgrund der unterschiedlichen pharmakophoren Eigenschaften werden an anderer Stelle mit der Bindetasche H-Brücken-Bindungen ausgebildet (vgl. Abb. 15.2a), weshalb die Liganden eine deutlich andere Orientierung in der Bindetasche aufweisen (vgl. Abb. 15.2b)¹.

Aus den unterschiedlichen Ligand-Rezeptor-Interaktionen von PDB 1dhf und PDB 1dls (s. Abb. 15.3) resultieren unterschiedliche PLIFs (s. Abb.15.4)¹. Der für die PLIFs erhaltene *Tanimoto-Koeffizient*= 0.50 weist deutlich auf die unterschiedlichen Bindungsmodi hin. Diese über topologische Fingerprints der Liganden zu erkennen, wäre nicht möglich gewesen.

¹Unter Verwendung von *Protonate-3D*²²² in den Standardeinstellungen wird MTX nicht protoniert, wie dies in NMR-Studien bestätigt worden ist¹. Ferner werden bei den PLIF-Berechnungen in MOE 2010.10 keine über Wassermoleküle vermittelten Ligand-Rezeptor-Interaktionen berücksichtigt⁶².

Die Bits für die kationische Interaktion bei PDB 1dls sowie zwei über Wassermoleküle vermittelte H-Brücken-Interaktionen bei PDB 1dhf und PDB 1dls (vgl. Abb. 15.2a) sind somit in den PLIFs nicht gesetzt worden. Weitere Details über Ligand-Rezeptor-Interaktionen von DHF und MTX mit der DHFR gibt KLEBE¹.

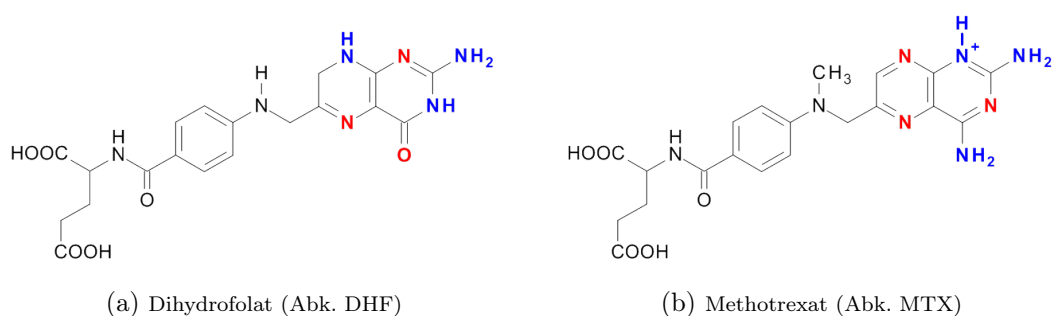
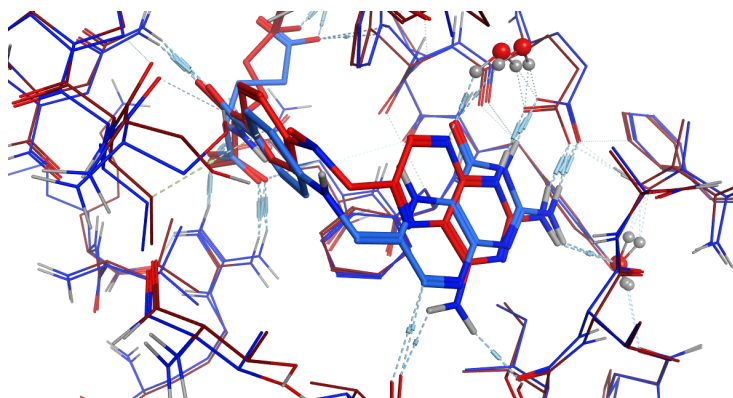
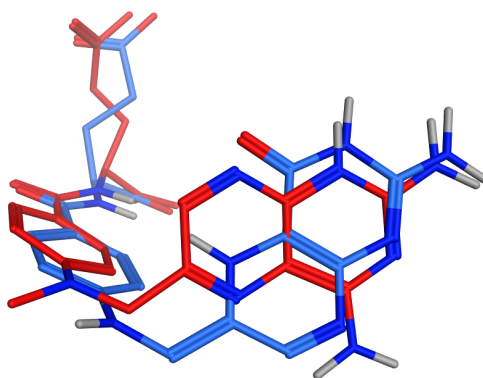


Abbildung 15.1: Pharmakophore Eigenschaften von DHF und MTX, die eine H-Brückenbindung mit der DHFR eingehen sind farblich hervorgehoben. Aufgrund der unterschiedlichen Anordnung von H-Brücken-Donoren und -Akzeptoren bei DHF und MTX ordnen sich die Liganden unterschiedlich in der Bindetasche an. *Blau: H-Brücken-Donor; rot: H-Brücken-Akzeptor*. Abbildung verändert nach KLEBE¹. Weitere Erläuterungen s. Text.



(a) Überlagerung der Bindetaschen von PDB 1dhf (blau; Ligand: DHF; Protein: DHFR) und PDB 1dls (rot; Ligand: MTX; Protein: DHFR) mit unterschiedlichen Ligand-Rezeptor-Interaktionen (gepunktete Linien).



(b) Orientierung von DHF (blau) und MTX (rot) in der Bindetasche nach Protein-Überlagerung.

Abbildung 15.2: Darstellung der unterschiedlichen Bindungsmodi von DHF (PDB 1dhf; blau) und MTX (PDB 1dls; rot) nach Überlagerung und Entfernung der Bindetasche (Darstellungen in MOE).

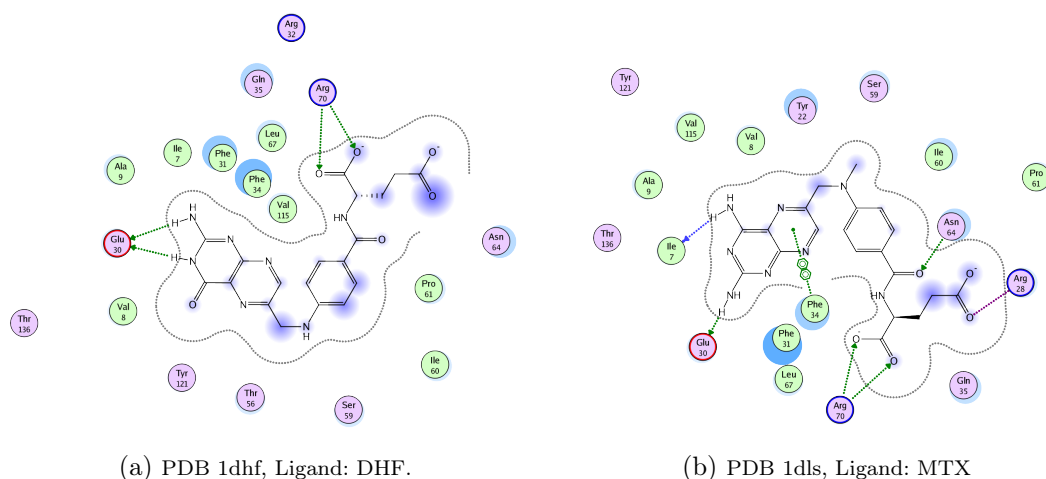


Abbildung 15.3: 2D-Darstellung der unterschiedlichen Ligand-Rezeptor-Interaktionen von DHF und MTX mit der DHFR (Darstellung in MOE).

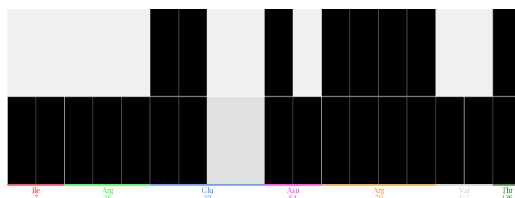


Abbildung 15.4: Die unterschiedlichen Ligand-Rezeptor-Interaktionen bei PDB 1dhf (DHF; oben) und PDB 1dls (MTX; unten), die zu deutlich unterschiedlichen PLIFs führen ($Tanimoto-Koeffizient = 0.50$; Darstellung in MOE).

15.2 Parallel Screening

Anstatt Pharmakophor-Suchen nur mit einem einzigen Modell durchzuführen, verwendeten STEINDL et al. in ihrer *Parallel Screening* Methode^{280–282} unterschiedliche Ligand-Rezeptor-Komplexe von viralen Zielstrukturen, die als Basis für strukturbasierte Pharmakophor-Ableitungen dienten. Die Autoren waren mit den parallel verwendeten Pharmakophor-Modellen in der Lage, Aktivitäten von antiviralen Substanzen an unterschiedlichen Zielstrukturen in einem retrospektiven Screening vorherzusagen (engl. *activity profiling*).

Die Grundidee des *Parallel Screenings* ist in dieser Arbeit aufgegriffen worden, um das Anreicherungsverhalten von Datenbanksuchen zu verbessern, im Vergleich zu Screenings, die nur ein Modell verwenden.

15.3 PLIF Driven Parallel Screening (PLIDriPaS)

Am Beispiel der *Phosphodiesterase-5* (Abk. PDE5) soll gezeigt werden, wie sich Screening-Ergebnisse durch die Kombination von PLIFs und *Parallel Screening* erheblich verbessern lassen. PDE5 ist als Beispiel gewählt worden, da in diesem Datensatz die strukturelle Variabilität zum einen am größten ist (26 unterschiedliche Liganden; 22 unterschiedliche reduzierte Graphen-Cluster, s. Tab. 14.1) und zum anderen mehrere unterschiedliche Bindungsmodi bekannt sind⁵⁶.

15.3.1 Datenvorbehandlung PLIDriPaS

Von allen in der *Screening-PDB*²³¹ verfügbaren PDE5-Ligand-Rezeptor-Komplexen (zehn Ligand-Rezeptor-Komplexe; Stand: 24.06.2011) wurde der vierstellige PDB-Code verwendet und die Strukturen entsprechend aus der Protein-Datenbank^{160,161} heruntergeladen^{II}. Die Ligand-Rezeptor-Komplexe wurden mit der MOE⁶² *Protonate-3D*-Methode²²² vorbehandelt. Kristallisationsartefakte, Gegen-Ionen, Co-Faktoren und Wassermoleküle^{III} wurden aus den Ligand-Rezeptor-Komplexen gelöscht. Von jedem Protein wurde jeweils ein Monomer betrachtet und die restlichen Ketten gelöscht. Die Nummerierung der Aminosäure-Reste wurde über ein .svl Programm (*resIDnfs.svl*; erhältlich unter²⁸⁸) standardisiert. Die so vorbehandelten Ligand-Rezeptor-Komplexe wurde in MOE überlagert und schließlich die PLIFs berechnet. Abbildung 15.5 zeigt die zehn superpositionierten Ligand-Rezeptor-Komplexe (oben links), die aus der Bindetaschen isolierten Liganden, die eine unterschiedliche Orientierung annehmen (oben rechts) und die auf Grundlage der Komplexe berechneten PLIFs (unten mitte).

^{II}Die in der sc-PDB hinterlegten Röntgenkristall-Strukturen erfüllen gewisse Qualitätsstandards, z.B. bzgl. der Auflösung. Ferner sind keine Liganden hinterlegt, die kovalent mit ihrer Zielstruktur interagieren²³¹. Die Ligand-Rezeptor-Komplexe sind nicht direkt aus der sc-PDB verwendet worden, da diese Strukturen bereits prozessiert worden sind. Um Verzerrungen der Screening-Ergebnisse entgegen zu wirken, sind die Röntgenkristall-Strukturen aus der PDB verwendet und eigenhändig aufgearbeitet worden.

^{III}Über Wassermoleküle vermittelte Ligand-Rezeptor-Interaktionen werden im PLIF (MOE 2010.10) nicht berücksichtigt, weshalb die Wassermoleküle gelöscht worden sind.

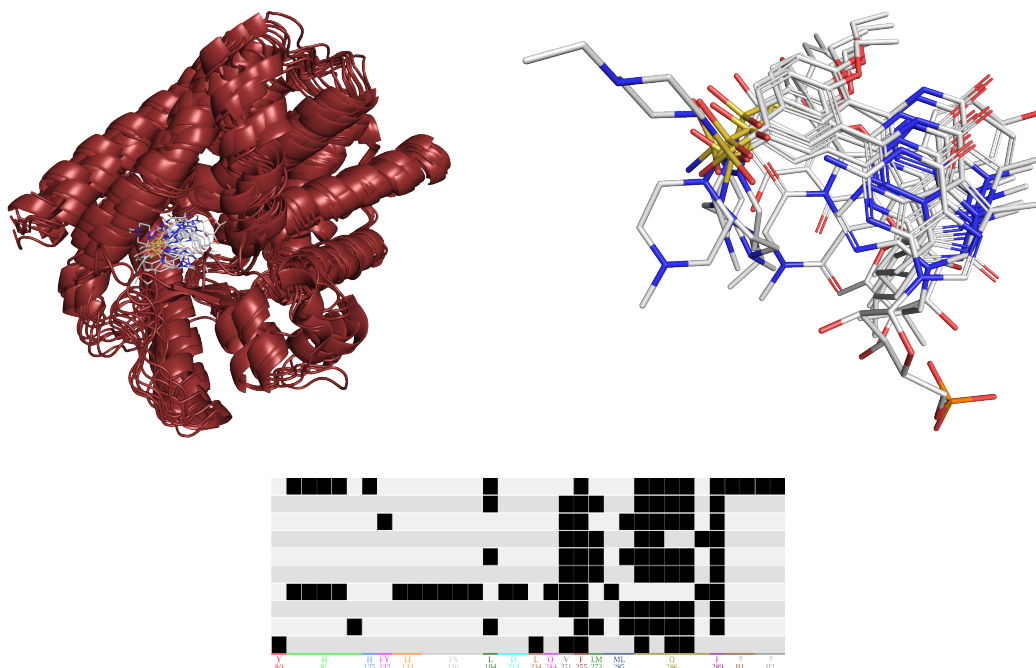


Abbildung 15.5: Zehn superpositionierte Phosphodiesterase-5-Ligand-Rezeptor-Komplexe (oben links). Die Liganden binden in unterschiedlichen Modi (oben rechts), was deutlich an den unterschiedlichen PLIF-Mustern zu erkennen ist (unten mitte).

15.3.2 Modellauswahl und Datenbanksuche

Für die Erstellung der Pharmakophor-Modelle sollten gezielt Komplexe verwendet werden, bei denen die co-kristallisierten Liganden alternative Bindungsmodi zeigen (deren PLIFs sich also stark unterscheiden), um einen größeren chemischen Raum abzudecken. Die erhaltenen PLIFs wurden visuell inspiziert und verwandte Interaktionsmuster jeweils zusammen gruppiert (engl. *clustering*)^{IV}.

Aus jedem PLIF-Cluster wurde jeweils ein Ligand-Rezeptor-Komplex zufällig ausgewählt. Auf dieser Grundlage wurde mit CavKA je ein Pharmakophor-Modell für jeden Ligand-Rezeptor-Komplex erstellt, das im MOE .ph4 Format (CCG PCHD Pharmakophor-Schema von 2010) gespeichert wurde. Die Modelle wurden anschließend optimiert und in einem virtuellen Screening genutzt. Als Programm für die Datenbanksuche wurde MOE verwendet. Bei der Modell-Optimierung wurde versucht, möglichst spezifische Modelle zu erhalten (mit einer möglichst kurzen Hitliste).

^{IV}Eine Gruppierung von SIFts nach visueller Inspektion wurde von DENG et al. durchgeführt²⁸³. Auch GRIMSHAW empfiehlt nachdrücklich ein solches Vorgehen für PLIFs und rät von *hierarchischem*- oder *k-means-Clustering* zur Gruppierung ähnlicher PLIFs ab²⁸⁹.

Die erhaltenen Hitlisten, stellvertretend für jeden Cluster, wurden vereinigt. Duplikate wurden aus den Hitlisten entfernt, sodass jedes Molekül nur einmal in der vereinigten Hitliste erscheint. Die für die Pharmakophor-Ableitung verwendeten Ligand-Rezeptor-Komplexe, die berechneten PLIFS, die Cluster-Zugehörigkeit sowie die Werte für *Sensitivität* und *Spezifität* der optimierten Pharmakophor-Modelle werden im Ergebnisteil in Kapitel 20.1 diskutiert.

Kapitel 16

CavKA: Versuchsaufbau

In den vorherigen Abschnitten dieser Arbeit sind die unterschiedlichen Techniken näher erläutert worden, bzw. ist generell auf die Validierung von Pharmakophor-Modellen eingegangen worden - nicht jedoch auf die Qualität, bzw. Validität der *Methode CavKA* selbst.

Als Arbeitshypothese ist davon ausgegangen worden, dass eine 'bessere' Methode in der Lage sein muss, Schlüssel-Interaktionen zuverlässiger zu erkennen und somit Modelle zu generieren, die eine höhere Anreicherung aktiver Moleküle zeigen. Eine detaillierte Übersicht, wie CavKA gegen andere Methoden zur Pharmakophor-Ableitung verglichen worden ist, folgt in diesem Kapitel.

16.1 Versuchsaufbau: Manuelle Zusammenführung geometrischer und energetischer Information

Um die Leistungsfähigkeit der nach Abschnitt 13.1 erstellten CavKA-Pharmakophor-Modelle zu beurteilen, ist anstatt eines absoluten Gütemaßes der Vergleich gegen Ligandscout, als etablierte Methode zur automatischen Ableitung strukturbasierter Pharmakophor-Modelle auf Basis von Ligand-Rezeptor-Komplexen, gewählt worden. Für beide Methoden wurden hierbei ausschließlich Harte-Sphären-Pharmakophor-Modelle verwendet. Für die nach Abschnitt 14.5.1 aufbereiteten Ligand-Rezeptor-Komplexe wurden mittels CavKA (unter Zuhilfenahme von MIFs) und Ligandscout auf Basis von Ligand-Rezeptor-Komplexen Pharmakophor-Modelle erstellt, die anschließend optimiert wurden. Für die gleichen Eigenschaften wurden in CavKA und Ligandscout identische Sphären-Durchmesser definiert. Den Modellen wurden keine weiteren Ligandeigenschaften hinzugefügt, die nicht von der jeweiligen Methode erkannt worden sind.

Die Hypothesen sind in das MOE .ph4 Format exportiert^I (hierbei wurde das CCG PCHD Pharmakophor-Schema von 2010 verwendet) und MOE als Software-Plattform für die Datenbanksuchen genutzt worden. Schließlich wurden die manuell optimierten Modelle über die *Cluster gewichtete Sensitivität* (vgl. Gl. 14.7) und *Spezifität* (vgl. Gl. 14.2) bewertet, wobei spezifische Modelle mit einer kurzen Hitliste bevorzugt worden sind.

Das oben beschriebene Vorgehen verdeutlicht, wie viel Aufwand des Anwenders nötig ist, bzw. man kann sich leicht vorstellen, wie subjektiv der Prozess der Pharmakophor-Optimierung sein kann. Dies hat zur Folge, dass ein auf dieser Basis durchgeführter Vergleich niemals völlig objektiv sein kann. Darüber hinaus bedeutet der Vergleich von zwei manuell optimierten Modellen nicht, dass die Existenz von Modellen mit einem besseren Anreicherungsvermögen ausgeschlossen werden kann. Keine gebräuchliche Methode zur Pharmakophor-Ableitung wird 'auf Knopfdruck' ideale Modelle liefern⁷⁶. Es muss stets eine Optimierung folgen, die grundsätzlich die Gefahr einer zu starken Anpassung an die spezifischen Gegebenheiten eines Datensatzes birgt und somit im prospektiven Screening nicht gleichermaßen erfolgreich sein wird. Dieses Phänomen nennt man im Englischen auch *overfitting*.

16.2 Versuchsaufbau: MIF-basierte E-Pharmakophore

Die in Abschnitt 13.2 definierten Regeln wurden verwendet, um MIF-basierte E-Pharmakophore für fünf der 13 Fieldscreen-Zielstrukturen zu erstellen. Die Pharmakophor-Modelle wurden im MOE .ph4 Format abgespeichert (hierbei wurde das CCG PCHD Pharmakophor-Schema von 2010 verwendet). Für die Datenbanksuche wurde MOE genutzt. Aufgrund der Eindeutigkeit der Regeln zur Erstellung der Hypothesen bezüglich essentieller Eigenschaften und partiell erfüllter Modelle im Rahmen der E-Pharmakophor-Methode¹⁶⁴ sind die Hypothesen bewusst nicht weiter optimiert worden.

Die von SALAM et al.¹⁶⁴ publizierten Pharmakophore für die PDB Ligand-Rezeptor-Komplexe 1cx2 (COX2), 1eve (AChE), und 1m17 (EGFR), 1rt1 (HIV-RT) und 1kv2 (P38) wurden in MOE rekonstruiert und mit den MIF-basierten E-Pharmakophoren verglichen.

^ILigandscout (Version 3.0.3) nutzt ein sehr viel älteres .ph4 Format von 2004, das z.T. stark fehlerbehaftet ist. Die exportierten Modelle wurden in MOE inspiziert, ggf. manuell korrigiert und im aktuellen CCG PCHD Pharmakophor-Schema von 2010 als .ph4 Datei gespeichert.

Ferner erfolgte ein Vergleich gegen die CavKA-, bzw. Ligandscout-Modelle mit der höchsten Anreicherung, um einen Eindruck zu erhalten, wie leistungsfähig rein energiebasierte Pharmakophor-Modelle im Vergleich zu CavKA-, bzw. Ligandscout-Modellen sind.

Die fünf oben genannten Ligand-Rezeptor-Komplexe sind gewählt worden, da sie sowohl in DUD als auch von SALAM et al. als Referenz verwendet worden sind¹⁶⁴.

16.3 Versuchsaufbau: CavKA^{HYBRID}

Bei den mit CavKA^{HYBRID} (vgl. Abschn. 13.3) durchgeführten Versuchen sollte analysiert werden, inwieweit sich Pharmakophor-Modelle ohne Anwenderintervention ableiten lassen und wie die Vorteile strukturbasierter Pharmakophore mit ligandzentrischen Ansätzen verknüpft werden können. Ferner wurde untersucht, welchen Einfluss die Repräsentation der verbotenen Sphären (Gauß-, bzw. Harte-Sphären) auf die Screening-Ergebnisse hat. Die Leistungsfähigkeit dieser Hybrid-Modelle wurde mit der rein ligandzentrischen Pharao-Methode verglichen. Für alle erstellten Modelle wurde Pharao als Programm zur Datenbanksuche verwendet²⁰³. Bei allen abgeleiteten Pharmakophor-Modellen wurden pharmakophore Ligandeigenschaften über die Pharao-Methode definiert¹⁸⁷. Interagierende Ligandeigenschaften (Modelle 3. und 4.) wurden durch das geometrische CavKA-Modul abgeleitet (vgl. Kapitel 11). Es wurde keine Kraftfeldinformation beim Erstellen der Hypothesen berücksichtigt. Insgesamt wurden für jeden der 13 DUD-Referenz Ligand-Rezeptor-Komplexe fünf Pharmakophor-Modelle nach dem folgenden Muster erstellt:

1. Für den co-kristallisierten Liganden wurde ein ligandzentrisches Pharao-Pharmakophor-Modell erstellt (**alle** Ligandeigenschaften wurden als Gauß-Sphären definiert; Abk. *ligandzentrisch*).
2. Für den co-kristallisierten Liganden wurde ein ligandzentrisches Pharmakophor-Modell erstellt (**alle** Ligandeigenschaften wurden als Gauß-Sphären definiert). Zusätzlich wurde der Rezeptor durch verbotene Sphären (ebenfalls Gauß-Sphären) repräsentiert (Abk. *ligandzentrisch + Gaussian*).
3. Für den Ligand-Rezeptor-Komplex wurde mit CavKA ein strukturbasiertes Pharmakophor-Modell erstellt (alle mit dem Rezeptor **interagierenden** Ligandeigenschaften wurden dabei als Gauß-Sphären definiert). Zusätzlich wurde der Rezeptor durch verbotene Sphären (ebenfalls Gauß-Sphären) repräsentiert (Abk. *CavKA + Gaussian*).

4. Für den co-kristallisierten Liganden wurde ein ligandzentrisches Pharmakophor-Modell erstellt (**alle** Ligandeigenschaften wurden als Gauß-Sphären definiert), auf das die Datenbank-Moleküle superpositioniert wurden. Im nächsten Schritt wurden die superpositionierten Moleküle gegen die verbotenen Harten-Sphären abgeglichen, die den Rezeptor repräsentierten (Abk. *ligandzentrisch + HYBRID*).
5. Für den Ligand-Rezeptor-Komplex wurde mit CavKA ein strukturbasiertes Pharmakophor-Modell erstellt (alle mit dem Rezeptor **interagierenden** Ligandeigenschaften wurden dabei als Gauß-Sphären definiert), auf das die Datenbank-Moleküle superpositioniert wurden. Im nächsten Schritt wurden die superpositionierten Moleküle gegen die verbotenen Harten-Sphären abgeglichen, die den Rezeptor repräsentierten (Abk. *CavKA + HYBRID*).

Bei den geschilderten Ansätzen stellen die Ansätze 4. und 5. die eigentlichen Hybrid-Ansätze dar, bei denen Gauß-Sphären und Harte-Sphären kombiniert werden (vgl. Abschn.13.3). Für den Strafterm $Tan_{Penalty}$ ist empirisch ein Wert von -0.7 festgelegt worden. Für Tol_{Dist} ist bewusst ein sehr großzügiger Wert von 0.5 Å definiert worden, was eine starke Annäherung des superpositionierten Moleküls an die verbotenen Sphären erlaubt.

Der niedrige Wert für Tol_{Dist} trug der groben Konformer-Auflösung von maximal 100 Konformeren pro Molekül Rechnung und sollte einer erhöhten *falsch-negativ* Rate entgegenwirken.

Ziel dieser Untersuchungen war es nicht, die bestmögliche Anreicherung zu erzielen, sondern vielmehr den Einfluss des jeweiligen Modells auf die Screening-Ergebnisse zu untersuchen. Dies setzt voraus, dass zumindest eine gewisse Anreicherung aktiver Moleküle zu beobachten sein muss. Da dies unter Verwendung der ursprünglichen Fieldscreen-Referenz Ligand-Rezeptor-Komplexe nicht immer gegeben war, wurden alternative Screening-Referenzen gesucht, beispielsweise durch ein *Data-Mining* in der sc-PDB²³¹. Alternativ wurde auf bereits publizierte Referenzen anderer Studien zurückgegriffen (s. Tab. 16.1).

Für alle durchgeführten Datenbanksuchen wurde Pharao²⁰³ als Programm in den Standardeinstellungen verwendet.

Zu beachten ist, dass bei den strukturbasierten Pharmakophor-Modellen 2. und 3. der Rezeptor durch verbotene Gauß-Sphären repräsentiert wird: Bei einer Datenbanksuche wird zunächst die Überlappung von Modell und Datenbank-Molekül maximiert und die verbotenen Sphären erst im Anschluss mit dem optimierten Modell abgeglichen, wobei das Datenbank-Molekül gegebenenfalls abgewertet wird. Alternativ dazu können verbotene Sphären bei jedem Optimierungsschritt berücksichtigt werden.

Zugunsten zuverlässigerer Ergebnisse wird empfohlen, die Standardeinstellungen zu verwenden. Werden verbotene Sphären bei jedem Optimierungsschritt berücksichtigt, besteht die Gefahr, dass die Prozedur der Superpositionierung in einem lokalen Minimum zwischen zwei verbotenen Sphären verbleibt. Dies führt letztlich zu suboptimalen Superpositionierungen und verfälschten Hitlisten. Pharaos zeigt sich besonders empfindlich gegenüber dieser Problematik, wenn sehr viele verbotene Sphären gesetzt worden sind²⁹⁰.

Tabelle 16.1: Alternative Ligand-Reptor-Komplexe als Referenzen für den Fieldscreen-Datensatz. Bei den mit *gekennzeichneten Strukturen wurde (mangels alternativer Strukturen, oder aufgrund bereits gutem Anreicherungsverhalten) der Fieldscreen-Referenz Ligand-Rezeptor-Komplex verwendet.

Zielstruktur	PDB ID	Auflösung (Å)	Referenz
ACE	1o86*	2.00	184
AChE	1eve*	2.50	164,184
CDK2	1e1x	2.20	231
COX2	1cx2*	3.00	164,184
EGFR	1m17*	2.60	164,184
FXa	1lpz	2.40	291
HIV-RT	2rf2	2.40	231
INHA	1p44*	2.70	184
P38	1a9u	2.50	164
PDE5	1tbf	1.30	231
PDGFRB	1t46*	1.60	184,291
SRC	2src*	1.50	184
VEGFR2	1fgi*	2.50	184

Kapitel 17

Versuchsaufbau und Ziele: ligandzentrischer Ansätze

Um die Leistungsfähigkeit Pharaos beurteilen zu können, erfolgte ein systematischer Vergleich mit anderen ligandzentrischen Methoden. Kernfragestellungen hierbei waren zum einen, mit welchen Methoden sich pharmakophore Ligandeigenschaften am besten modellieren lassen (Volumen, elektrostatische Potentiale oder durch regelbasierte Definition), zum anderen war es von Interesse, wie gut die Superpositionierung, bzw. die Optimierung der jeweiligen Methode geeignet ist, um eine Anreicherung aktiver Moleküle in einem retrospektiven Screening zu erzielen. Zusätzlich sollte untersucht werden, wie leistungsfähig die frei verfügbaren *Open-Source*-Methoden im Vergleich zu den etablierten, kommerziell erhältlichen zu beurteilen sind. So wurden die ligandzentrischen Methoden ROCS, EON, Pharaos, Pyramid und ShaEP (s. Kap. 7) unter Verwendung unterschiedlicher Bewertungsfunktionen für ein Virtuelles Screening verwendet.

Auch wenn die Fieldscreen-Referenz-Liganden häufig nur begrenzt repräsentativ für die Menge an aktiven Molekülen sind, wurden sie dennoch verwendet, um eine Vergleichbarkeit mit anderen Methoden zu ermöglichen, die ebenfalls den Fieldscreen-Validier-Datensatz verwenden.

Teil III

Ergebnisse und Diskussion

Kapitel 18

Ergebnisse und Diskussion: CavKA versus Ligandscout

Abbildung 18.1 zeigt die erzielten Leistungskennzahlen *Sensitivität* und *Spezifität* der manuell optimierten Pharmakophor-Modelle, die mit CavKA und Ligandscout erstellt worden sind. Als Validier-Datensatz wurde der nach Abschnitt 14.4.2.1 vorbehandelte Fieldscreen-Datensatz verwendet (struct_HTS). Die Ligand-Rezeptor-Komplexe, die als Basis für die Erstellung der Pharmakophor-Modelle verwendet wurden, sind gemäß Abschnitt 14.5.1 vorbehandelt worden. Als Programm zur Datenbanksuche diente MOE.

Beim Betrachten der Ergebnisse ist zu beobachten:

- CavKA zeigt auf den Zielstrukturen ACE, AChE, CDK2, EGFR, PDE5 und PDGFRB gegenüber Ligandscout ein besseres Anreicherungsverhalten über Zufallsniveau.
- Ligandscout zeigt auf den Zielstrukturen COX2 und INHA gegenüber CavKA ein besseres Anreicherungsverhalten über Zufallsniveau.
- Für die Zielstrukturen FXa und VEGFR2 konnten mit CavKA und Ligandscout nur Modelle erstellt werden, deren Anreicherungsverhalten sich auf Zufallsniveau bewegt. Auch für die Zielstrukturen INHA und HIV-RT gelang es unter Verwendung der CavKA-Methode nur Modelle zu erstellen, die eine Anreicherung auf Zufallsniveau aufwiesen. Indes wurden für die mit Ligandscout für die Zielstrukturen CDK2 und SRC erstellten Modelle ein schlechteres Anreicherungsverhalten beobachtet, als es bei zufälliger Selektion von Molekülen zu erwarten gewesen wäre.

-
- Für die Zielstruktur SRC war es mit CavKA, für die Zielstrukturen HIV-RT und PDGFRB mit Ligandscout, nicht möglich, sinnvolle Modelle zu erstellen (*Sensitivität*= 0, *Spezifität*= 0; *Sensitivität*= 1, *Spezifität*= 0; *Sensitivität*= 0, *Spezifität*= 1). Modelle mit derartigen Kennzahlen führen dazu, dass eine, bzw. beide der Hitlisten für Aktive und / oder Köder-Moleküle keinen Treffer aufweist. Eine Separation von Aktiven und Köder-Molekülen gelingt nicht, weshalb solche Modelle keinen Nutzen im Rahmen eines Virtuellen Screenings bieten.
 - Für Pharmakophor-Modelle, wie z.B. bei P38 (CavKA) gilt, dass trotz eines nominellen Anreicherungsverhaltens über Zufallsniveau die Spezifität sehr niedrig ist. Die erhaltenen Hitlisten sind somit sehr groß, was die Testkapazitäten (v.a. im akademischen Bereich) häufig deutlich überschreitet. Ob sich somit aus diesen Pharmakophor-Modellen eine praktische Relevanz ergibt, hängt sehr stark von den verfügbaren Assay-Kapazitäten ab.

Das bessere Abschneiden der CavKA-Methode lässt sich in vielen Fällen auf die Interaktionserkennung hydrophober Ligandeigenschaften zurückführen. Tendenziell ist CavKA sensibler und detektiert mehr interagierende Ligandeigenschaften als Ligandscout. Die mit CavKA erstellten Pharmakophore bieten somit mehr Variationsmöglichkeiten für die Kombination pharmakophorer Eigenschaften, womit die Wahrscheinlichkeit steigt, ein Modell mit einem besseren Anreicherungsverhalten zu erstellen. Nachdrücklich ist noch einmal zu betonen, dass für den Vergleich zwischen CavKA und Ligandscout manuell optimierte Modelle verwendet wurden. Möglicherweise existieren andere Kombinationen pharmakophorer Eigenschaften mit einem besseren Anreicherungsverhalten (z.B. durch die Berücksichtigung projizierter pharmakophorer Ligandeigenschaften).

Nachfolgend wird das Anreicherungsverhalten der mit CavKA, bzw. Ligandscout erstellten Pharmakophor-Modelle für eine Auswahl der Zielstrukturen des Fieldscreen-Datensatzes diskutiert. Dabei wird aufgezeigt, inwiefern MIFs zur Erstellung von Pharmakophor-Modellen im Rahmen der CavKA-Methode genutzt werden können. Mit CavKA erstellte Modelle zeigen in den meisten Fällen ein besseres Anreicherungsverhalten als Ligandscout-Modelle, die nach rein geometrischen Kriterien erstellt worden sind. Bei allen erzielten Ergebnissen ist zu berücksichtigen, dass sie nur auf Basis eines einzigen Ligand-Rezeptor-Komplexes erstellt worden sind, der im Allgemeinen einen nur sehr begrenzten Raum für das Fenster an möglichen Interaktionen bereitstellt.

Ist die Menge der aktiven Moleküle sehr heterogen bzgl. ihrer Grundgerüste, bzw. binden aktive Moleküle in unterschiedlichen Bindungsmodi, ist die erzielte Anreicherung meist nur leicht über Zufallsniveau. Aus diesem Grund wird später die PLIDriPaS-Methode evaluiert, da hier verschiedene Bindungsmodi berücksichtigt werden können.

Der Einfluss der verwendeten Referenz, bzw. der Einfluss der Konformeren-Auflösung auf die Screening-Ergebnisse wird in den Abschnitten 18.2 und 18.3 diskutiert. Eine tabellarische Zusammenfassung der in diesem Kapitel diskutierten Ergebnisse des Vergleichs von CavKA mit Ligandscout findet sich im Anhang in Abschnitt 28.1. Aufgeführte Dateinamen bezeichnen Pharmakophor-Modelle, die in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie der Technischen Universität Braunschweig archiviert sind.

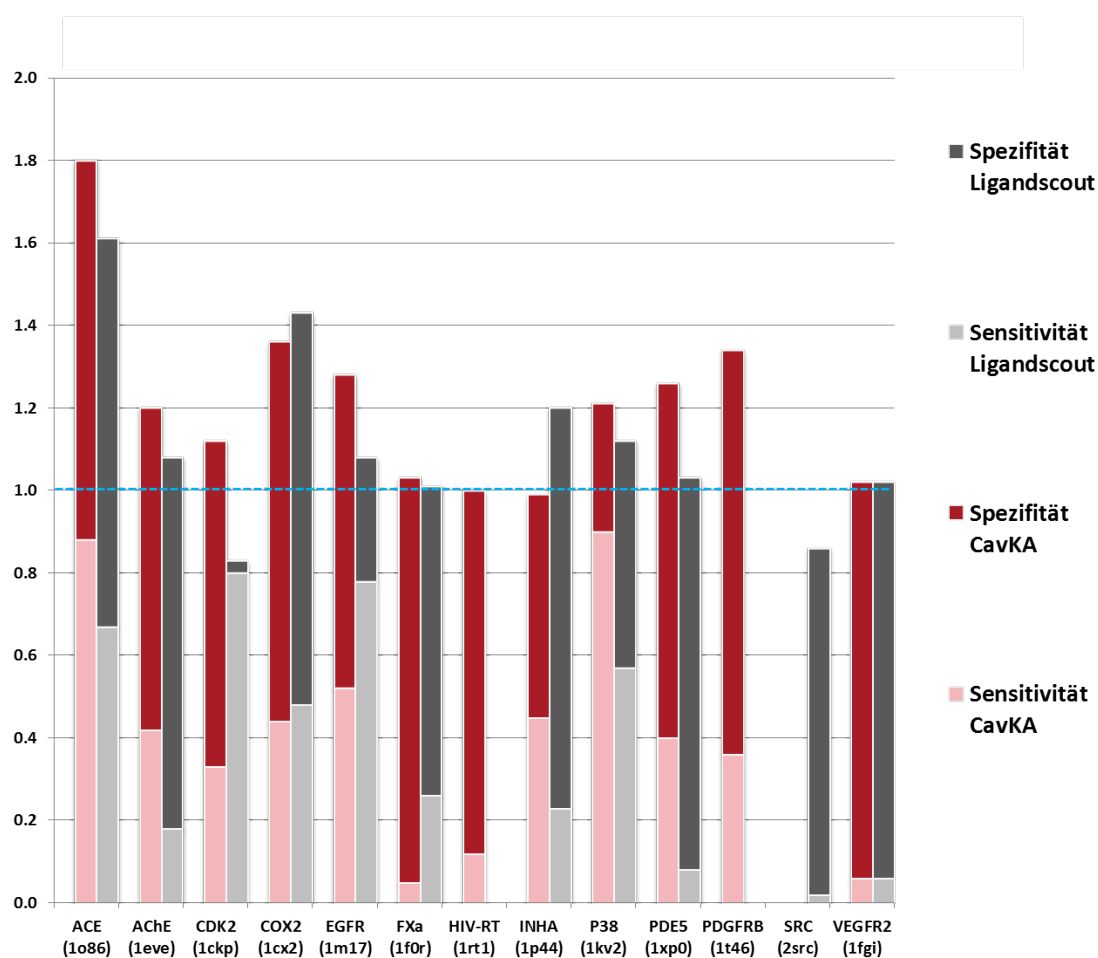


Abbildung 18.1: *Sensitivität* und *Spezifität* der manuell optimierten Pharmakophor-Modelle von CavKA (rot) und Ligandscout (grau) für die 13 Fieldscreen-Zielstrukturen (struct_HTS). Blau gepunktete Linie: Zufallsniveau.

18.1 CavKA: Vorteile durch MIF Information

18.1.1 CDK2: Identifikation hydrophober Interaktionen

Cyklinabhängige-Kinasen (engl. *cyclin dependent kinases*; Abk. CDK) spielen eine wichtige Rolle bei der Regulation des Zellzyklus, der Apoptose, der Transkription, der Zelldifferenzierung sowie bei der Regulation von Vorgängen im zentralen Nervensystem²⁹². CDK-Inhibitoren werden daher unter anderem als Therapeutika bei malignen Tumoren verwendet, um dysregulierte Zellzyklen beeinflussen zu können^{292–294}.

Bei dem für die Modell-Erstellung verwendeten DUD-Referenz Ligand-Rezeptor-Komplex der *Cyklinabhängige-Kinase 2* (Abk. CDK2; PDB 1ckp) handelt es sich um einen Liganden, der an die aktivierte Form der Kinase bindet (engl. *DFG-in binder*; auch Typ I-Inhibitor genannt). Typ I-Inhibitoren blockieren die Substratbindestelle für das native Substrat Adenosintriphosphat (Abk. ATP; vgl. Abb. 18.2a)¹.

Abbildung 18.2b zeigt den Pharmakophor für Typ I-Inhibitoren, der unter anderem durch eine hydrophobe Interaktion mit der sogenannten Adenin-Region gekennzeichnet ist. Darüber hinaus wird / werden zwischen eine(r) und drei H-Brücken-Interaktion(en) zur Scharnierregion (engl. *hinge region*) ausgebildet^{1,295}. Die Interaktion zur *hinge region* wird von nahezu allen Kinase-Inhibitoren ausgebildet¹.

Abbildung 18.3 zeigt die mit Ligandscout und CavKA identifizierten Ligand-Rezeptor-Interaktionen sowie die auf dieser Basis erstellten Pharmakophor-Modelle. Ligandscout erkennt bei der Ligand-Rezeptor-Analyse kaum hydrophobe Interaktionen in der Adenin-Region (vgl. Abb. 18.3a) und erstellt ein sehr unspezifisches Dreipunkt-Pharmakophor (vgl. Abb. 18.3b), das für die Praxis keine Relevanz besitzt (*Sensitivität* = 0.80; *Spezifität* = 0.17).

Im Gegensatz dazu detektiert CavKA die hydrophobe Interaktion zwischen dem Purin-Grundgerüst und der Adenin-Region nach geometrischen Kriterien (grüne Sphären in Abb. 18.3c). Die Relevanz, der nach geometrischen Kriterien ermittelten Eigenschaften, wird durch ein konturiertes, hydrophobes C3-MIF bestätigt (grüne Kreuze in Abb. 18.3c, die sich mit den grünen Sphären decken). Die hydrophobe Region des *DFG-in* Pharmakophors wird klar erkannt. Hieraus resultiert ein deutlich spezifischeres Fünfpunkt-Pharmakophor (vgl. Abb. 18.3d) mit einem klar besseren Anreicherungsverhalten (*Sensitivität* = 0.33; *Spezifität* = 0.79).

Bei dem co-kristallisierten Liganden aus dem Referenz Ligand-Rezeptor-Komplex handelt es sich um ein recht kleines Molekül, das den *DFG-in* Pharmakophor quasi mit Minimalanforderungen erfüllt. Aufgrund der geringen Ligandgröße ist die Anzahl möglicher Ligand-Rezeptor-Interaktionen stark eingeschränkt, und somit sind auch die Kombinationsmöglichkeiten pharmakophorer Eigenschaften begrenzt¹. Mit größeren Liganden, die ein größeres Fenster für die Kombination pharmakophorer Eigenschaften bieten, ist ein besseres Anreicherungsverhalten zu erwarten.

¹Die in der PDB hinterlegte Kristallstruktur ist fehlerhaft. Es handelt sich bei dem co-kristallisierten Liganden nicht wie von GRAY et al.²⁹⁶ in der Originalpublikation behauptet um Purvanalol B.

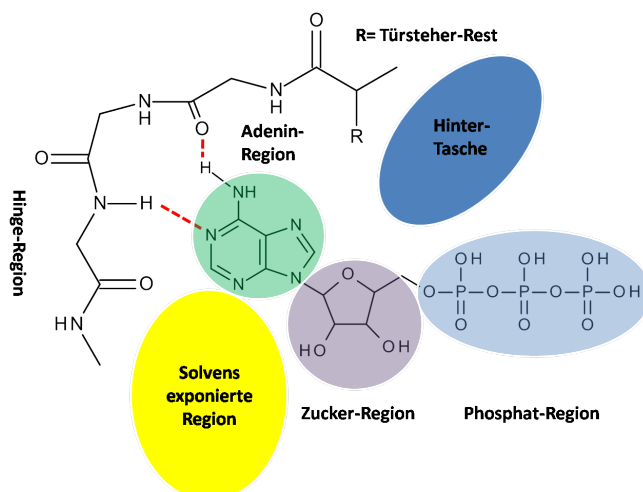
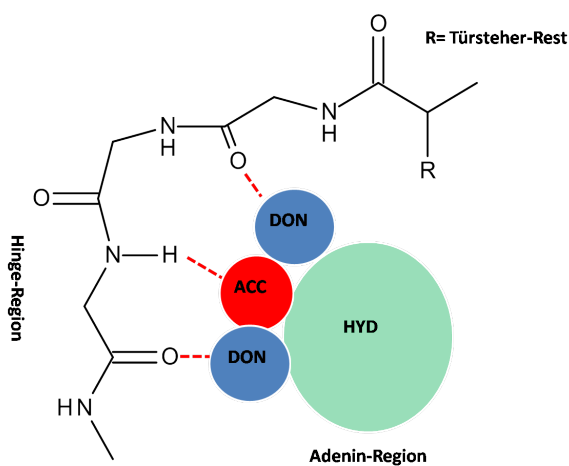
(a) Sogenanntes Traxler-Modell²⁹⁷ der ATP-Bindetasche.(b) Pharmakophor für Typ I-Kinase-Inhibitoren²⁹⁵.

Abbildung 18.2: Schematischer Überblick über die ATP-Bindestelle in Kinasen sowie der Pharmakophor von Typ I-Kinase-Inhibitoren. Weitere Erläuterungen s. Text. *Rot-gestrichelte Linie: Interaktion zur Hinge-Region; HYD: hydrophob; ACC: H-Brücken-Akzeptor; DON: H-Brücken-Donor.* (Abb. nach ZUCCOTTO et al.²⁹⁵).

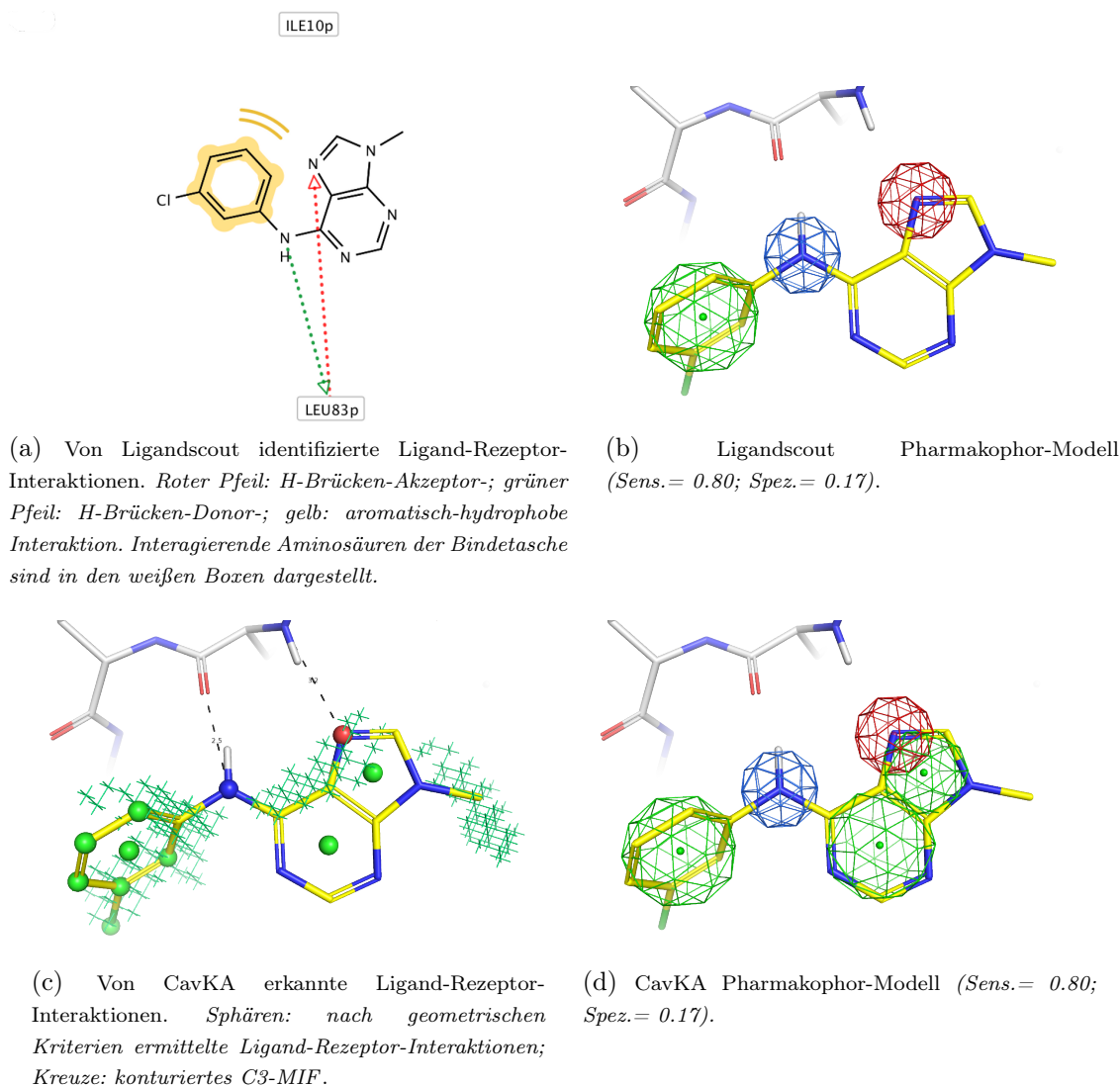


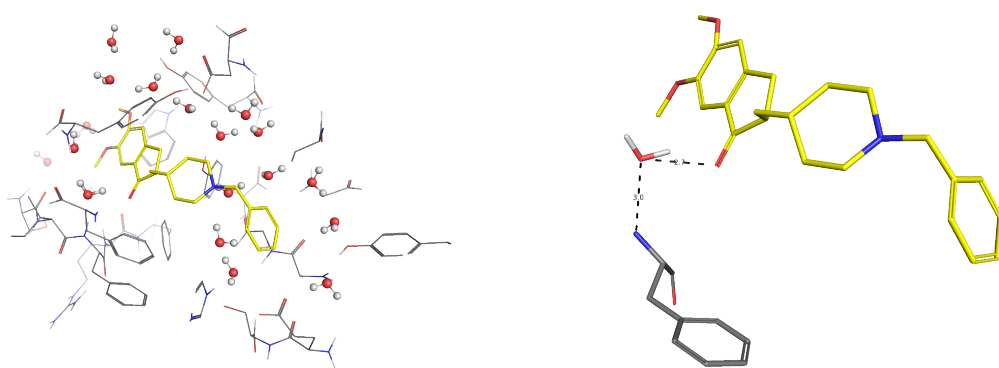
Abbildung 18.3: Vergleich der hydrophoben Ligand-Rezeptor-Interaktionserkennung zwischen Ligandscout und CavKA sowie abgeleiteter Pharmakophor-Modelle am Beispiel der *Cyklinabhängigen-Kinase 2* (PDB 1ckp). Blau: H-Brücken-Donor Eigenschaften; rot: H-Brücken-Akzeptor Eigenschaften; grün: hydrophobe Eigenschaften. Sens.: Sensitivität; Spez.: Spezifität. Verbotene Sphären übersichtshalber nicht gezeigt.

Am Beispiel der CDK2 wird deutlich, dass hydrophobe Ligand-Rezeptor-Interaktionen mit geometrischen Kriterien allein nicht zuverlässig erkannt werden können (s. Ligandscout-Pharmakophor-Modell). CavKA verwendet sehr einfache und sensitive geometrische Kriterien zur Identifikation hydrophober Ligand-Rezeptor-Interaktionen (vgl. Abschn. 11.2), was dazu führt, dass tendenziell sehr viele hydrophobe Ligandeigenschaften als 'interagierend' erkannt werden. Eventuell *falsch-positive* hydrophobe Ligandeigenschaften werden durch den MIF-Abgleich wiederum reduziert, sodass relevante hydrophobe Interaktionen sicher identifiziert werden.

18.1.2 AChE: Identifikation wichtiger Wassermoleküle

Acetylcholinesterase (Abk. AChE) stellt eine wichtige Zielstruktur bei der Behandlung von Demenzerkrankungen dar. Inhibitoren der Acetylcholinesterase verhindern den Abbau des Neurotransmitters Acetylcholin zu Acetat und Cholin. Eine Erhöhung des Acetylcholinspiegels geht häufig mit der Verbesserung kognitiver Fähigkeiten einher²⁹⁸.

Bei dem DUD-Referenz Ligand-Rezeptor-Komplex handelt es sich bei dem co-kristallisierten Liganden um das Antidementivum Donepezil (Aricept®). Wichtige Ligand-Rezeptor-Interaktionen wie hydrophobe, aromatische π - π -Interaktionen und Kationen- π -Interaktionen, wie sie bereits von KRYGER et al. beschrieben worden sind²⁹⁹, werden von CavKA zuverlässig erkannt. Eine Interaktion zwischen dem Dimethoxyindanon-Rest und Phe 288 findet nicht unmittelbar zwischen Ligand und Bindetasche statt, sondern wird über eines der zahlreichen co-kristallisierten Wassermoleküle (HOH 1254) vermittelt (vgl. Abb. 18.4a und Abb. 18.4b)²⁹⁹. Bei der Erstellung eines Pharmakophor-Modells ergibt sich die Möglichkeit, die Akzeptoreigenschaft des Dimethoxyindanon-Rests in der Pharmakophor-Hypothese zu berücksichtigen (Modell A^{II}; s. Abb. 18.5a) oder diese auszulassen (Modell B^{III}; s. Abb. 18.5b). Mit Modell A (*Sensitivität*= 0.42; *Spezifität*= 0.78) werden gegenüber Modell B deutlich bessere Anreicherungen erzielt (*Sensitivität*= 0.42; *Spezifität*= 0.67). Durch das Berücksichtigen der Akzeptor-Eigenschaft erhöht sich die *Spezifität* des Modells dramatisch, wohingegen die *Sensitivität* unverändert bleibt, was klar die Relevanz dieser Interaktion belegt.



(a) Co-kristallisierte Wassermoleküle in der Bindetasche.

(b) Über HOH 1254 vermittelte Ligand-Rezeptor-Interaktion.

Abbildung 18.4: Co-kristallisierte Wassermoleküle in dem Ligand-Rezeptor-Komplex PDB 1eve (*Acetylcholinesterase*; AChE). Weitere Erläuterungen s. Text.

^{II}Dateiname: 8_ache_1eve_struct_based_CCG_i_WATER.ph4

^{III}Dateiname: 5_ache_1eve_struct_based_CCG_i.ph4

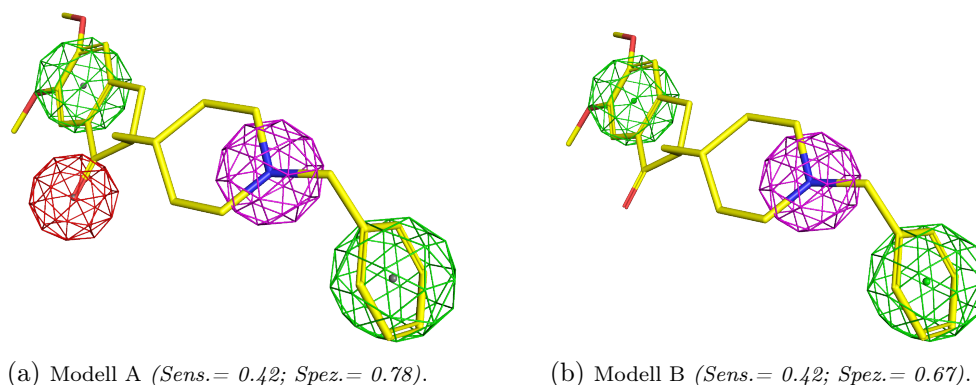


Abbildung 18.5: Mit CavKA auf Basis von PDB 1eve (*Acetylcholinesterase*; Abk. AChE) erstellte Pharmakophor-Modelle. Weitere Erläuterungen s. Text. *Grün*: aromatische; *magenta*: kationische; *rot*: H-Brücken-Akzeptor-Eigenschaften. Verbotene Sphären übersichtshalber nicht gezeigt.

In diesem Zusammenhang bietet sich ein weiterer Verwendungszweck für MIFs an. Sie können genutzt werden, um aus der Vielzahl von co-kristallisierten Wassermolekülen, deren Relevanz nicht unmittelbar ersichtlich ist, Schlüssel-Wassermoleküle zu bestätigen. Abbildung 18.6 zeigt die Übereinstimmung des relevanten Wassermoleküls mit einem konturierten MIF der GRID-Wassersonde H₂O. Darüber hinaus können H₂O-MIFs nicht nur zur zusätzlichen Absicherung verwendet werden, sondern auch, wenn keine Wassermoleküle co-kristallisiert sind. So könnten potentiell wichtige Wassermoleküle vorhergesagt werden, die Ligand-Rezeptor-Interaktionen vermitteln.

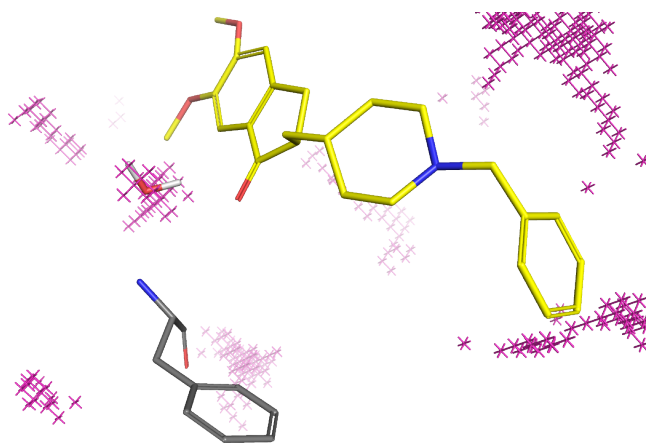


Abbildung 18.6: Bestätigung des Wassermoleküls HOH1254 durch ein konturiertes H₂O-MIF (magenta).

Es ist zu betonen, dass auch Ligandscout die Schlüssel-Interaktion des Wassermoleküls über geometrische Kriterien zuverlässig erkennt. Dies setzt jedoch das Vorhandensein co-kristallographierter Wassermoleküle voraus.

Darüber hinaus werden die Kationen- π -Interaktion und die 'Face-to-Face'- π - π -Interaktion des Benzyl-Rests korrekt von Ligandscout identifiziert. Die Leistungsfähigkeit des CavKA-Modells wird dennoch nicht erreicht, da eine hydrophobe Interaktionen zwischen dem Dimethoxyindanon-Rest und der Bindetasche nicht erkannt wird. (s. Abb. 18.7)^{IV}.

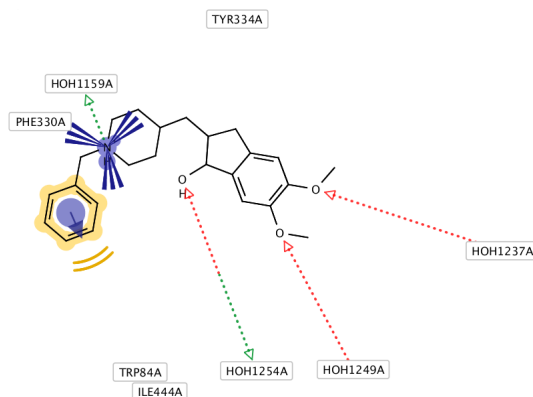


Abbildung 18.7: Von Ligandscout identifizierte Ligand-Rezeptor-Interaktionen. Erläuterungen s. Text. (PDB 1eve; Ligand: Donepezil). Rote Pfeile: H-Brücken-Akzeptor-; grüner Pfeil: H-Brücken-Donor-; gelb: aromatisch-hydrophobe-; blauer Kreis mit Pfeil: 'Face-to-Face'- π - π -; blauer Stickstoff: Kationen- π -Interaktion(en). Interagierende Aminosäuren und Wassermoleküle der Bindetasche sind in den weißen Boxen dargestellt.

18.1.3 PDGFRB: Priorisierung pharmakophorer Eigenschaften

Bei der *Platelet Derived Growth Factor Receptor Kinase Beta* (Abk. PDGFRB) handelt es sich ebenfalls um eine Kinase, die eine Zielstruktur für die Behandlung myeloischer Erkrankungen darstellt³⁰⁰.

Bei dem co-kristallisierten Referenz-Liganden Imatinib (Glivec[®]) aus dem Referenz Ligand-Rezeptor-Komplex PDB 1t46 handelt es sich um ein großes Molekül, das an die inaktive Form der Kinase bindet (sog. *DFG-out binder*; auch Typ II-Inhibitor genannt)³⁰¹. Abbildung 18.8a zeigt, dass in diesem Beispiel nahezu jede Ligandeigenschaft mit der Bindetasche interagiert. Das unmittelbar mit CavKA nur nach geometrischen Kriterien abgeleitete Pharmakophor-Modell ist in Abbildung 18.8b dargestellt. Es besteht aus 13 pharmakophoren Sphären, die eine Menge an Kombinationsmöglichkeiten bieten, um ein sinnvolles Pharmakophor-Modell zu erstellen.

^{IV}Die Carbonylfunktion des Ligand-Rezeptor-Komplexes im .pdb Format wird von Ligandscout (auch nach manueller Aufarbeitung) fälschlicherweise als Hydroxylfunktion erkannt. In MOE hingegen erfolgt eine korrekte Darstellung.

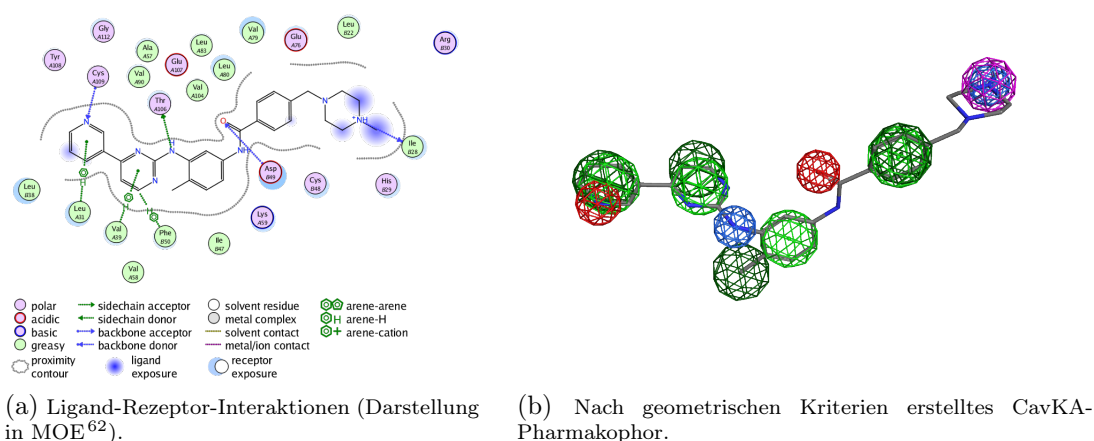


Abbildung 18.8: Nach geometrischen Kriterien erstelltes CavKA-Pharmakophor-Modell von PDGFRB (PDB 1t46). Rote Sphären: H-Brücken-Akzeptor; blaue Sphären: H-Brücken-Donor; grüne Sphären: hydrophobe; violette Sphären: kationische Eigenschaften. Verbotene Sphären übersichtshalber nicht gezeigt.

Bei derartig vielen Eigenschaften kann die MIF-Information als eine erste Hilfestellung verwendet werden, um eine Priorisierung pharmakophorer Eigenschaften für die Modell-Erstellung vorzunehmen. So kann beispielsweise eine erste Vorauswahl pharmakophorer Sphären getroffen werden. Ferner können *partial-* und *must-matches* über die *Hotspot*-Information definiert werden. Abbildung 18.9 zeigt das exemplarische Vorgehen.

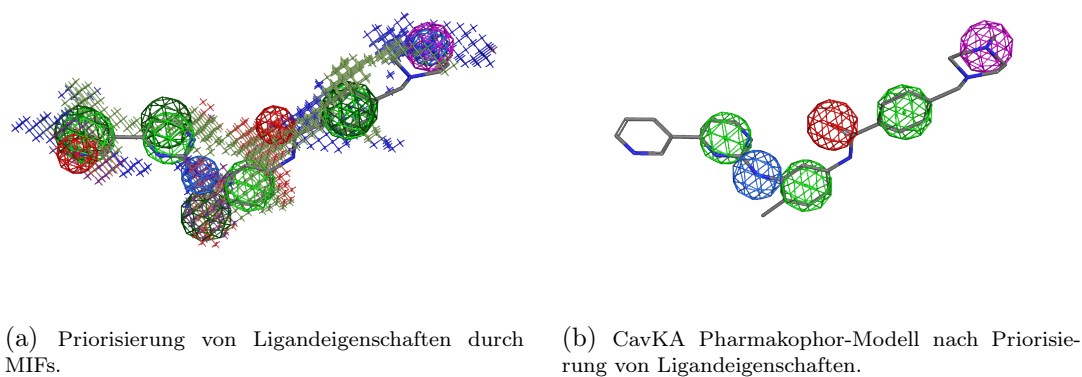


Abbildung 18.9: Nach Abgleich geometrischer Kriterien mit MIFs erstelltes Pharmakophor-Modell von PDGFRB (PDB 1t46). Rote Sphären: H-Brücken-Akzeptor; blaue Sphären: H-Brücken-Donor; grüne Sphären: hydrophobe; violette Sphären: kationische Eigenschaften. Rote Kreuze: O_MIF; Blaue Kreuze: N1_MIF; grüne Kreuze: C3_MIF. Verbotene Sphäre übersichtshalber nicht gezeigt.

Auch wenn bei einem solchen Vorgehen immer noch viel Einfluss des Anwenders gefordert ist (insbesondere *falsch-positive* MIF-Information erschwert die Modell-Erstellung), so lassen sich relevante Eigenschaften leichter erkennen und die ursprüngliche Anzahl pharmakophorer Eigenschaften reduzieren.

Obwohl das Pharmakophor-Modell in Abbildung 18.9 auf Basis einer Röntgenkristallstruktur der cKIT-Kinase anstatt des DUD-PDGFRB-Homologie-Modells erstellt worden ist (vgl. Abschn. 14.3.3), konnte eine *Sensitivität* von 0.36 und eine *Spezifität* von 0.98 erzielt werden. Dies mag aufgrund der verwendeten Referenz und der Tatsache, dass es sich bei den Aktiven um PDGFRB-Inhibitoren handelt, verwundern. Von Imatinib ist jedoch bekannt, dass es neben der cKIT-Kinase auch die Abl-Kinase und PDGFRB zu inhibieren vermag²⁹⁵. Dies deutet auf die engen Verwandtschaftsbeziehungen der Bindetaschen hin und erklärt letztlich die relativ hohe *Spezifität* bei einer guten *Sensitivität*.

18.2 Einfluss der Referenz auf die Screening-Ergebnisse

18.2.1 ACE

Inhibitoren des *Angiotensin-Konversionsenzyms* (engl. *angiotensin converting enzyme*; Abk. ACE), einer Zink-Protease, werden seit langer Zeit erfolgreich u.a. als Antihypertonika und zur Therapie der Herzinsuffizienz eingesetzt¹.

Bei ACE-Inhibitoren sind die Struktur-Wirkungsbeziehungen (und damit auch der Pharmakophor) klar beschrieben. Essentiell für die biologische Aktivität sind eine H-Brücken-Akzeptor-Eigenschaft sowie eine funktionelle Gruppe, die in der Lage ist, an das Zn^{2+} -Ion des aktiven Zentrums zu binden. Bei den therapeutisch eingesetzten ACE-Inhibitoren handelt es sich hierbei meist um eine Thiol- oder Carboxyl-Gruppe, bzw. Phosphinsäuren bei den moderneren Substanzen. Zusätzlich zu der Zink bindenden Funktionalität erwiesen sich Prolin-analoge Strukturen mit einer freien Carboxyl-Gruppe als stark aktivitätssteigernd. Bevorzugt werden Arzneistoffe mit einer Carboxyl-Gruppe verwendet, die weniger oxidationsempfindlich als Thiol-Gruppen haltige Arzneistoffe sind und eine längere Halbwertszeit besitzen. Zusätzlich zu den genannten Ligandeneigenschaften ist ein weiterer H-Brücken-Akzeptor für die biologische Aktivität essentiell¹. Abbildung 18.10a zeigt das mit CavKA erstellte Pharmakophor-Modell, welches ein enorm gutes Anreicherungsverhalten bewies (*Sensitivität* = 0.88; *Spezifität* = 0.92).

Abbildung 18.10b zeigt, wie interagierende Ligandeigenschaften darüber hinaus durch entsprechende MIFs bestätigt werden.

So ist das gute Anreicherungsverhalten letztlich darauf zurückzuführen, dass im verwendeten Fieldscreen-Referenz Ligand-Rezeptor-Komplex zwischen dem co-kristallisierten Liganden Lisinopril und der Bindetasche Ligand-Rezeptor-Interaktionen ausgebildet werden³⁰², die bei allen ACE-Inhibitoren anzutreffen sind¹ und diese von CavKA korrekt identifiziert werden. Das mit Ligandscout erstellte Modell schneidet indes schlechter ab (*Sensitivität* = 0.67; *Spezifität* = 0.94), da die hydrophobe Interaktion des Prolin-Rests mit der S'_2 Bindetasche nicht erkannt wird (s. Abb. 18.11).

Der Fieldscreen-ACE-Datensatz ist ein sehr positives Beispiel dafür, dass die verwendete Referenz repräsentativ für die Gesamtheit der aktiven Moleküle ist.

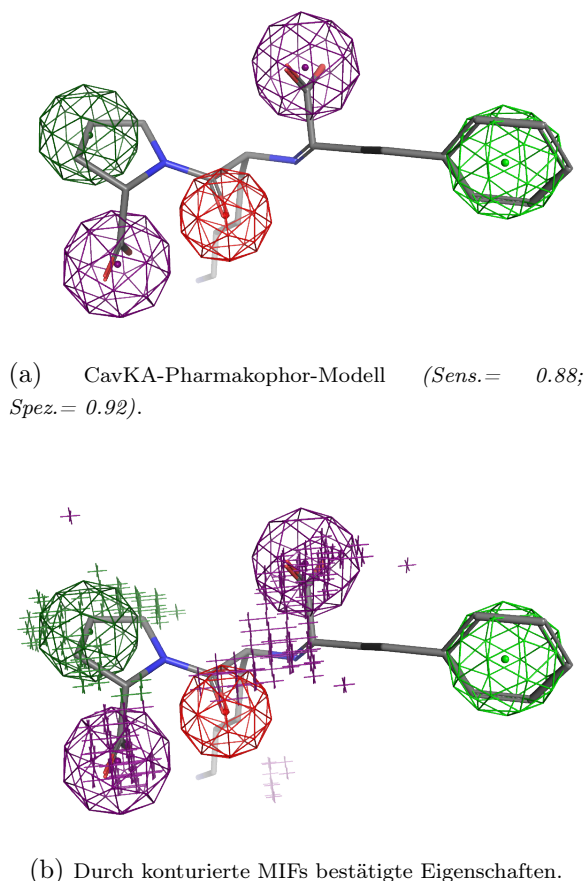


Abbildung 18.10: Auf Basis von PDB 1o86 (ACE; Ligand: Lisinopril) mit CavKA abgeleitetes Pharmakophor-Modell. *Purpur*: anionische; *hellgrün*: aromatische u. hydrophobe; *dunkelgrün*: hydrophobe; *rot*: H-Brücken-Akzeptor Eigenschaften. *Purpurne Kreuze*: konturiertes O^- -MIF; *dunkelgrüne Kreuze*: konturiertes DRY-MIF. *Sens.*: Sensitivität; *Spez.*: Spezifität. *O-MIF* übersichtshalber nicht gezeigt.

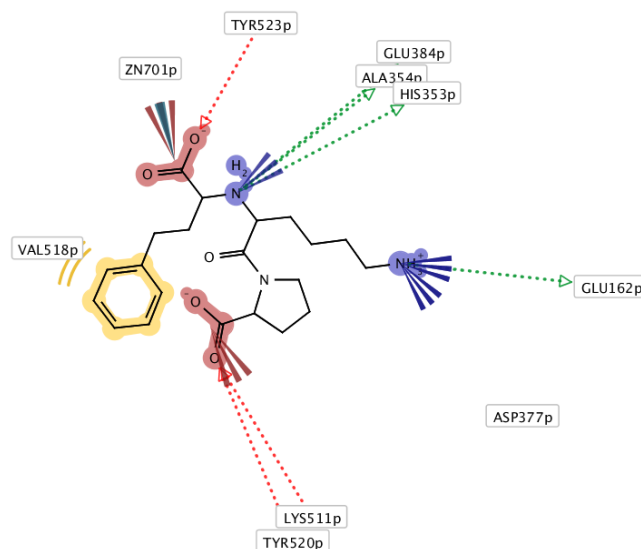


Abbildung 18.11: Von Ligandscout identifizierte Ligand-Rezeptor-Interaktionen (PDB 1o86; ACE; Ligand: Lisinopril). Die hydrophobe Interaktion zwischen Prolin-Rest und Bindetasche wird nicht erkannt. Rote Pfeile: H-Brücken-Akzeptor-; grüne Pfeile: H-Brücken-Donor-; gelb: aromatisch-hydrophobe-; blaue Stickstoffe: kationische-; braune Carboxylat-Gruppen: anionische Interaktion(en). Interagierende Aminosäuren der Bindetasche sind in den weißen Boxen dargestellt.

18.2.2 FXa

Faktor Xa (Abk. FXa) spielt eine wichtige Rolle in der Blutgerinnungskaskade und wird als Zielstruktur für Antithrombotika verwendet¹. FXa zählt zu einer sehr gut erforschten Zielstruktur, bei der zusätzlich zu einer großen Menge an Aktivitätsdaten viele Röntgen-Kristallstrukturen frei verfügbar sind^V.

Bei den FXa-Inhibitoren lassen sich grob zwei Strukturmerkmale identifizieren, die entscheidende Interaktionen mit dem Rezeptor eingehen und letztlich für die Inhibition verantwortlich sind¹. Hierbei handelt es sich in den meisten Fällen entweder um eine stark basische Funktionalität (sehr häufig einen Benzamidin-Rest) oder um eine hydrophobe Gruppe (häufig Chlorphenyl-, Chlornaphthyl- oder Chlorthiophen-Reste)¹. Bei den Inhibitoren, die eine basische Gruppe besitzen, kommt es zur Ausbildung einer sehr festen Salzbrücke mit dem anionischen Aspartat 189 in der FXa-Bindetasche¹. Die Vertreter mit einer hydrophoben Funktionalität sind in der Lage, ein stark konserviertes Wassermolekül am Boden der Bindetasche zu verdrängen, was mit einem enormen Zugewinn an Entropie einhergeht und die Bindung begünstigt³⁰³.

^VAllein in der sc-PDB²³¹ sind 68 qualitativ hochwertige Röntgenkristallstrukturen hinterlegt (Stand: 07.09.2011). In der PDB existieren darüber hinaus noch deutlich mehr Strukturen.

Die hydrophoben chlosubstituierten Derivate besitzen den Vorteil, dass sie ein besseres Selektivitätsprofil und bessere Bioverfügbarkeiten besitzen³⁰⁴. So erfolgte im Jahr 2008 die Markteinführung des ersten oral verfügbaren FXa-Inhibitors Rivaroxaban (Xarelto®), der ebenfalls einen hydrophoben Chlorthiophenyl-Rest besitzt¹. LEE und PLAYER geben in einer sehr umfangreichen Übersichtsarbeit weitere Einblicke über Grundstrukturmodifikationen und Aktivitätsdaten³⁰⁴.

Obwohl recht viel über die Zielstruktur bekannt ist, gelang es mit CavKA und Ligandscout auf Basis von PDB 1f0r nur Pharmakophor-Modelle zu erstellen, die aktive Moleküle auf Zufallsniveau anreicherten. Bei dem co-kristallisierten Liganden handelt es sich um einen Vertreter mit basischer Kopf-Gruppe.

Abbildung 18.12 zeigt die pharmakophoren Eigenschafts-Definitionen des co-kristallisierten Liganden aus dem vorbehandelten Referenz Ligand-Rezeptor-Komplex PDB 1f0r von MOE (Abb. 18.12a)^{VI}, Pharao^{VII} (Abb. 18.12b) und Ligandscout (Abb. 18.12c). Die Distanz zwischen der basischen Amino-Gruppe des Isochinolins und dem sauren Asp 189 ist mit 3.7 Å relativ groß (vgl. Abb. 18.12d), sodass es weder zu einer H-Brücken-Interaktion, noch zu einer Protonierung der Amino-Gruppe im Zuge der Vorbehandlung mit *Protonate-3D* kommt. Dies hat zur Folge, dass keine der drei Methoden eine kationische Eigenschaft definiert. In diesem speziellen Fall interagiert der Ligand mit dem Stickstoff des Isochinolins über ein Wassermolekül vermittelt mit dem Asp 189 der FXa-Bindetasche (vgl. Abb. 18.12d).

Von den 64 aktiven Molekülen des Fieldscreen-FXa-Datensatzes besitzen 60 Strukturen eine basische Funktionalität (93.75 %) und nur vier eine hydrophobe Funktionalität (6.25%). Der Grund für häufige Vorkommen der basischen Struktur ist historischer Natur. Erst Mitte der Neunziger Jahre fiel das Dogma, dass FXa-Inhibitoren eine basische Funktionalität besitzen *müssen*¹, weshalb deutlich weniger Inhibitoren mit einer hydrophoben funktionellen Gruppe existieren.

Aufgrund der nicht identifizierten kationischen Ligandeigenschaft ist die Referenz nur begrenzt mit den aktiven Molekülen vergleichbar. Wird bei der Definition von Ligandeigenschaften eine Eigenschaft nicht erkannt (in diesem Fall die kationische) kann im nächsten Schritt auch keine kationische Ligand-Rezeptor-Interaktion identifiziert werden. Eine kationische Eigenschaft fehlt folglich im Pharmakophor-Modell. Dieses Beispiel verdeutlicht: Fehler, die bei der Definition von Ligandeigenschaften gemacht werden, können nicht ohne weiteres automatisch korrigiert werden, was häufig zu fehlerhaften Pharmakophor-Modellen führt.

^{VI}Die MOE-Eigenschafts-Definitionen sind exemplarisch dargestellt worden, um eine weitere Vergleichsmöglichkeit aufzuführen.

^{VII}CavKA nutzt für die Definition pharmakophorer Ligandeigenschaften die Regeln der Pharao-Methode (vgl. Kap. 10.1).

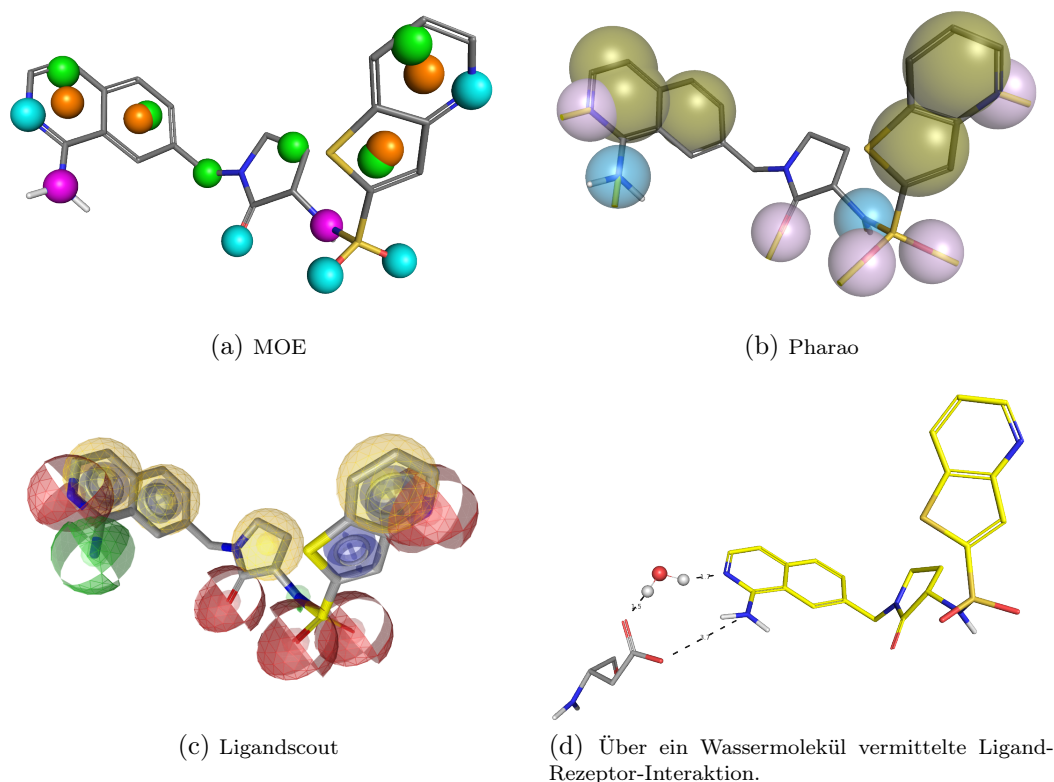


Abbildung 18.12: Eigenschafts-Definitionen von MOE, Pharao und Ligandscout des co-kristallisierten Liganden aus PDB 1f0r. Keine der Methoden definiert eine positive Eigenschaft für die basische Amino-Gruppe des Isochinolins (Erklärung siehe Text). *MOE-Eigenschaften*: türkis: Acc; orange: Aro; grün: Hyd; magenta: Don. *Pharao-Eigenschaften*: blau-HDON; flüchtig: HACC; oliv: HYBH. *Ligandscout Eigenschaften*: rot: Acc; grün: Don; gelb: Hyd; blau: Aro.

Fehlt die entscheidende kationische Ligandeigenschaft, die bei fast allen Aktiven vorhanden ist im erstellten Pharmakophor-Modell, ist es nicht verwunderlich, dass nur Anreicherungen auf Zufallsniveau erreicht werden.

Um diese Hypothese experimentell zu prüfen, wurde ein extrem reduziertes Pharmakophor-Modell mit nur einer aromatischen und einer kationischen Eigenschaft erstellt, ohne dabei Verbotene-Sphären zu definieren^{VIII}.

Mit diesem Modell wurde eine *Sensitivität* = 0.95 und eine *Spezifität* = 0.46 erzielt. Die Tatsache, dass mit so einfachen Modellen derartige Anreicherungen bei einer vermeintlichen Menge von diversen aktiven Molekülen erzielt worden sind, lassen den Fieldscreen-FXa-Datensatz für Validierungsexperimente nur begrenzt geeignet erscheinen.

^{VIII}Dateiname des Pharmakophor-Modells: 1_DUMMY.ph4; Datensatz-Vorbehandlung: struct_HTS.

18.2.3 P38

Die *P38 Mitogen aktivierte Protein Kinase* (Abk. P38) spielt eine wichtige Rolle bei der Regulierung der Produktion proinflammatorischer Cytokine wie Tumornekrosefaktor- α und Interleukin-1 β . Inhibitoren der P38 bieten somit eine Therapieoption bei vielen inflammatorischen Erkrankungen³⁰⁵.

Mit Pharmakophor-Suchen gelang es für die Zielstruktur P38 nur sehr unspezifische Pharmakophor-Modelle zu erstellen (*Sensitivität* CavKA= 0.90, *Spezifität* CavKA= 0.31^{IX}; *Sensitivität* Ligandscout= 0.57, *Spezifität* Ligandscout= 0.55^X). Beide Modelle weisen nur ein geringes Anreicherungsvermögen über dem Zufallsniveau auf. Um den Einfluss der *Methode* auf die Ergebnisse der Datenbanksuche zu analysieren, wurden eine Vielzahl unterschiedlicher ligandzentrischer Ansätze sowie das Docking-Programm FRED¹⁸⁹ für ein Virtuelles Screening auf dem P38-Datensatz verwendet. Überraschenderweise zeigte keine der verwendeten Methoden ein überzeugendes Anreicherungsverhalten (vgl. Abb. 18.13). Das lässt darauf schließen, dass die schlechten Ergebnisse nicht allein auf die verwendete Screening-Methode zurückzuführen sein können.

KINNINGS und JACKSON wiesen bereits auf das schlechte Anreicherungsvermögen hin, das beim Verwenden des P38-Referenz-Liganden aus PDB 1kv2 mit ihrer ligandzentrischen *Ligmatch*-Methode erzielt worden ist. Die Autoren begründeten dieses Phänomen damit, dass es sich bei BIRB 796, dem co-kristallisierten Liganden aus PDB 1kv2, zum einen um ein sehr großes Molekül (MW > 500) und zum anderen um einen wesentlich selteneren Typ II-Inhibitor handelt. Die Autoren empfahlen stattdessen den co-kristallisierten Liganden aus PDB 1bl7 als Alternativ-Referenz zu verwenden^{261 XI}.

Um den Einfluss der verwendeten Referenz auf die Screening-Ergebnisse zu analysieren, wurden alle in der sc-PDB²³¹ verfügbaren Ligand-Rezeptor-Komplexe der P38 heruntergeladen (68 Ligand-Rezeptor-Komplexe; Stand: 04.05.2011). Der jeweilige co-kristallisierte Ligand wurde für eine ROCS-Suche (Scoring-Funktion: *ROCS_{combo}*; Konformeren-Auflösung: HTS) verwendet. ROCS ist als Screening-Methode verwendet worden, da die Methode in der Regel gute Ergebnisse liefert.

^{IX}Datei: 6_P38_1kv2_struct_based_CCG_i.ph4

^XDatei: 12_P38_1kv2_struct_based_LIGANDSCOUT_PCHD.ph4

^{XI}Hierbei ist darauf hinzuweisen, dass unter Verwendung des co-kristallisierten Liganden aus PDB 1bl6 als Referenz für die ROCS-Suche sogar noch bessere Ergebnisse erzielt wurden als mit 1bl7. Abbildung 31.2 in Abschnitt 31.2 im Anhang zeigt die ROC-Kurven.

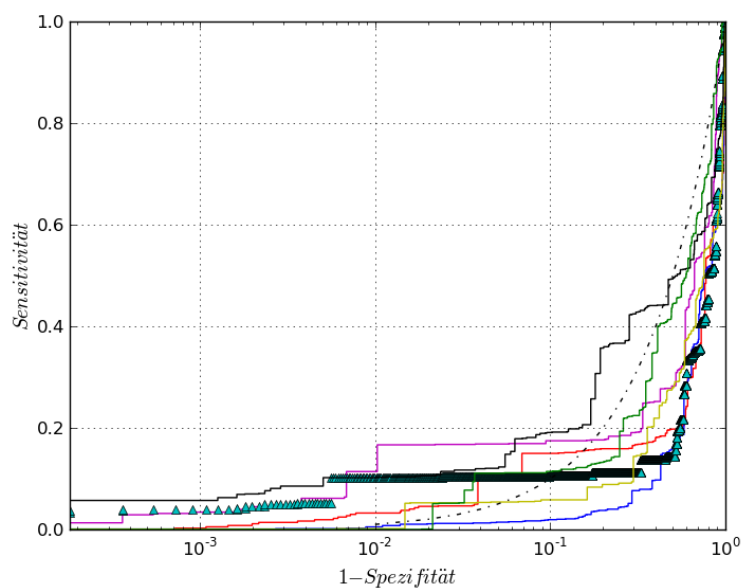


Abbildung 18.13: Ergebnisse eines Virtuellen Screenings mit unterschiedlichen Methoden auf dem Fieldscreen-P38-Datensatz (PDB 1kv2; ligcent_HTS). Rot: $ROCS_{combo}$; blau: $ROCS_{color}$; magenta: $ROCS_{shape}$; türkisarbenes Dreieck: Pharaos; gelb: Pyramid; schwarz: EON_{combo} ; grün: FRED; schwarz gepunktet: Zufall.

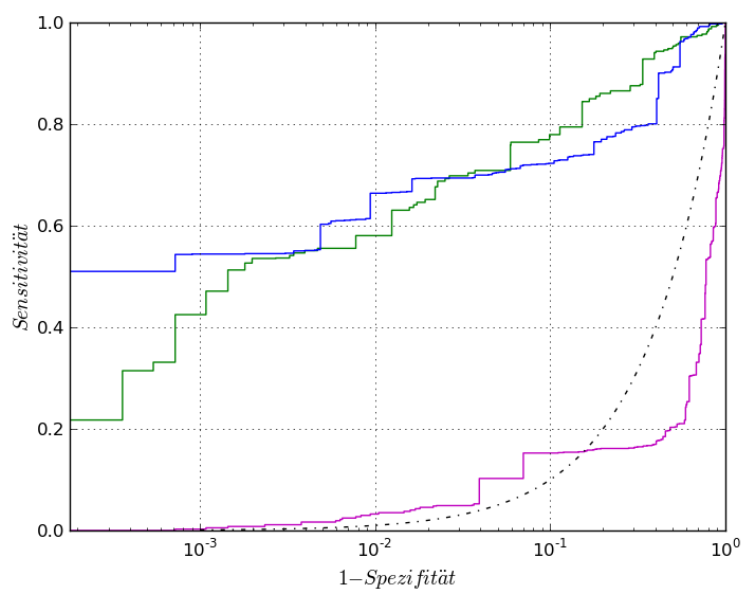
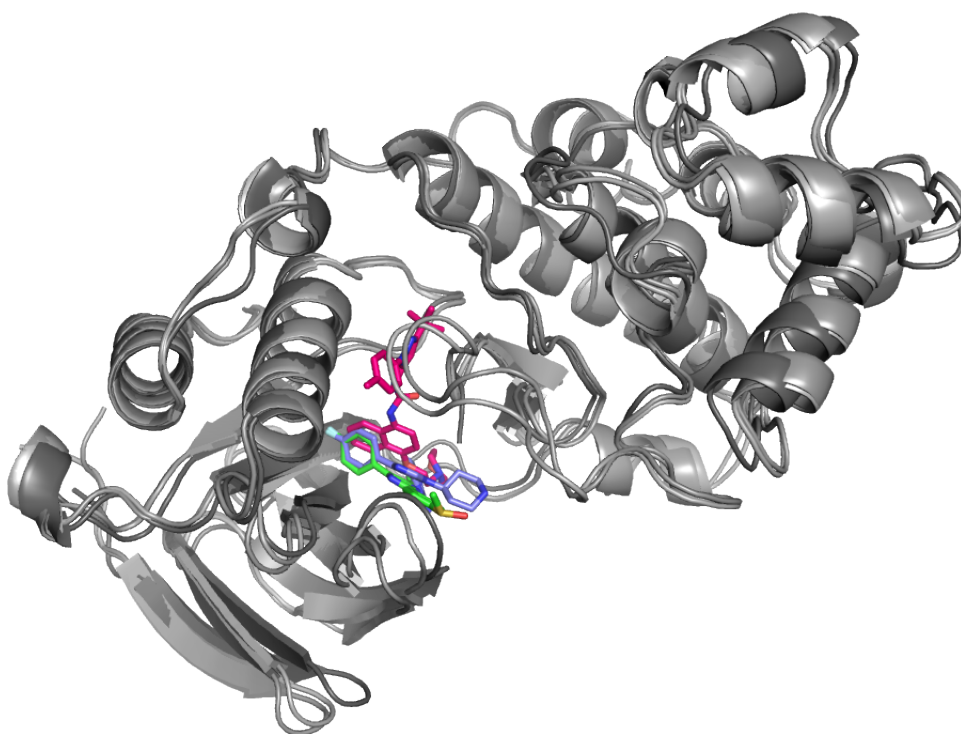


Abbildung 18.14: $ROCS_{combo}$ Screening-Ergebnisse auf dem Fieldscreen-P38-Datensatz (ligcent_HTS) unter Verwendung unterschiedlicher Referenz-Liganden. Grün: PDB 1a9u; blau: PDB 1bl7; magenta: PDB 1kv2; schwarz gepunktet: Zufall.

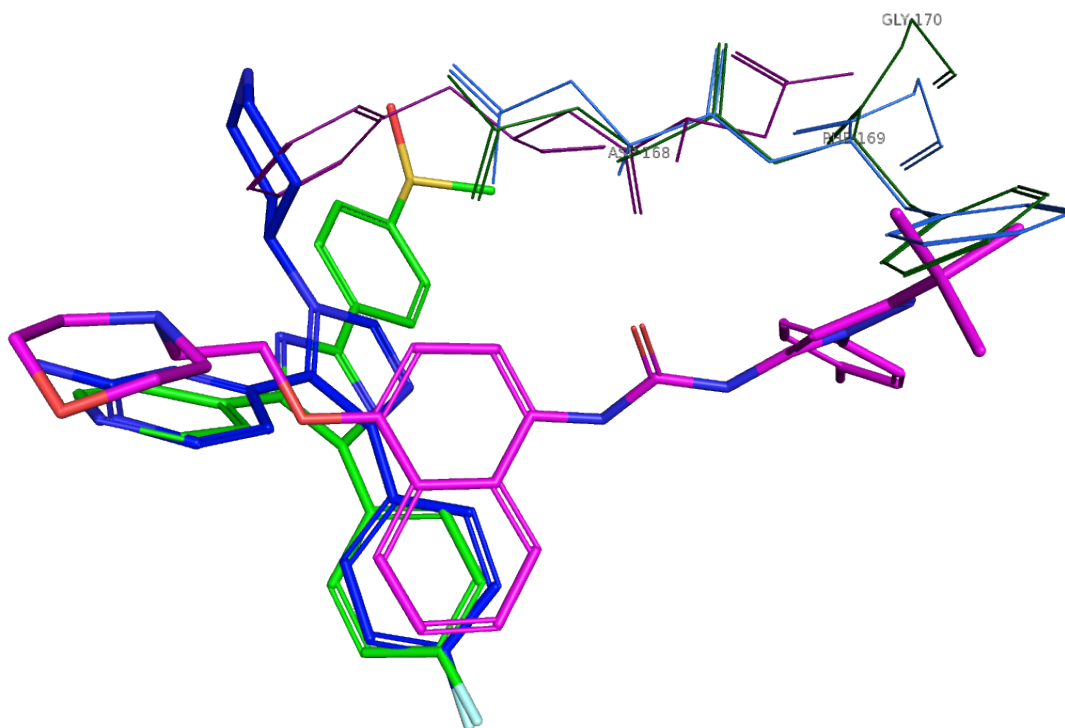
Exemplarisch sind in Abbildung 18.15b die ROCS-Screening-Ergebnisse für jene drei Liganden aus den PDB-Kristallstrukturen 1a9u (grün) und 1bl7 (blau) dargestellt, die ein sehr gutes Anreicherungsvermögen aufweisen. Ferner ist das Ergebnis, das mit dem Liganden der DUD-Referenz 1kv2 (magenta) erzielt worden ist, abgebildet. Letztgenannter beweist ein sehr schlechtes Anreicherungsvermögen aktiver Moleküle. Aus den erhaltenen Ergebnissen wird deutlich, dass die vermeintlich schlechten Ergebnisse, die unter Verwendung des Referenz-Liganden aus PDB 1kv2 erzielt worden sind, keinesfalls auf die Suchmethode ROCS zurückzuführen sind. Die Ergebnisse werden vielmehr sehr stark von der Auswahl des Referenz-Moleküls beeinflusst. Superpositioniert man die drei Bindetaschen (s. Abb. 18.15a) und betrachtet die Orientierung der Liganden in Raum, so wird ersichtlich, dass der magentafarbene Typ II-Inhibitor BIRB 796³⁰⁵ zum einen deutlich größer als die Alternativ-Referenzen mit gutem Anreicherungsverhalten ist und zum anderen eine deutlich andere Orientierung in der Bindetasche einnimmt (s. Abb. 18.15b). Bei den co-kristallisierten Liganden aus 1a9u und 1bl7 handelt es sich im Gegensatz zu BIRB 796 um Typ I-Inhibitoren. Die Moleküle binden am Türsteher-Rest vorbei und bilden zusätzlich in der Hinter-Tasche hydrophobe Ligand-Rezeptor-Interaktionen aus (engl. *back-pocket*; s. Abb. 18.2a)²⁹⁵.

Der Typ II-Inhibitor aus dem Referenz Ligand-Rezeptor-Komplex aus PDB 1kv2 ist offenbar nicht repräsentativ für die restlichen aktiven Moleküle des P38-Datensatzes. Ob es sich bei den Aktiven des P38-Datensatzes um Typ I- oder Typ II-Inhibitoren handelt, kann nicht in einem biochemischen Assay festgestellt werden²⁹⁵. Allein die Kristallstruktur kann hier Aufschluss bieten.

Zusammenfassend lässt sich schließen: So lange es sich bei den durchgeführten Analysen 'nur' um eine Modellvalidierung handelt, ist es eine zwingende Voraussetzung für die verwendete Referenz, dass sie repräsentativ für die Gesamtheit der aktiven Molekülen ist, um überhaupt eine Anreicherung erzielen zu können. Dies ist insbesondere bei der Fieldscreen-Referenz für P38 im Falle des Ligand-Rezeptor-Komplexes PDB 1kv2 nicht gegeben. Ergebnisse, die auf Basis solcher Screening-Referenzen erzielt werden, können keinerlei Rückschlüsse auf die Leistungsfähigkeit der verwendeten Methode geben.



(a) Die superpositionierten Ligand-Rezeptor-Komplexe 1a9u, 1bl7, 1kv2.



(b) Die Orientierung der Liganden und des *DFG*-Motivs nach Superpositionierung in der Bindetaschen. Die Typ I-Inhibitoren (grün und blau) binden an die aktive *DFG*-in Konformation der P38, der Typ II-Inhibitor (magenta) an die inaktive *DFG*-out Konformation. Grün: 1a9u; blau: 1bl7; magenta: 1kv2.

Abbildung 18.15: Unterschiedliche Bindungsmodi von P38-Inhibitoren.

18.3 Einfluss der Konformeren-Auflösung

Der nach Abschnitt 14.4.2.1 vorbehandelte Fieldscreen-Datensatz (`struct_HQS`) wurde für alle 13 Zielstrukturen mit einem manuell optimierten Pharmakophor-Modell durchsucht. Hierbei ist jenes Pharmakophor-Modell verwendet worden, mit dem die besten Ergebnisse auf der jeweiligen Zielstruktur mit niedrig aufgelösten Konformeren-Ensembles erzielt worden sind (`struct_HTS`). Bei dieser Betrachtung war die Methode, mit der die Pharmakophor-Modelle erstellt worden sind (Cavka, bzw. Ligandscout), sekundär: Es sollte einzig und allein der Einfluss der Konformeren-Auflösung auf die Screening-Ergebnisse analysiert werden. Die Methode, mit der das jeweilige Modell für die Datenbanksuche erstellt worden ist, ist im Anhang in Abschnitt 28.2 aufgeführt.

Abbildung 18.16 zeigt die Leistungskennzahlen *Sensitivität* und *Spezifität* in Bezug auf die CavKA- und Ligandscout-Pharmakophor-Modelle für die 13 Fieldscreen-Zielstrukturen unter Verwendung unterschiedlicher Konformeren-Auflösungen. Beim Betrachten der Ergebnisse wird deutlich, dass die Konformeren-Auflösung nur einen sehr geringen Einfluss auf die Screening-Ergebnisse hat.

Mit hochaufgelösten Konformeren-Ensembles (`struct_HQS`) werden indes nicht zwangsweise die besseren Ergebnisse erzielt. Es besteht stets die Gefahr, dass inaktive Köder-Moleküle den Pharmakophor erfüllen (was ggf. bei einer niedrigen Konformeren-Auflösung nicht der Fall gewesen wäre) und sich die *falsch-positiv* Rate erhöht (vgl. ACE, AChE und CDK2). Für die Zielstruktur HIV-RT lässt sich eine Zunahme der *Sensitivität* unter Verwendung des `struct_HQS`-Ensembles beobachten. Dennoch ist die zu beobachtende Anreicherung nur schwach über dem Zufallsniveau. Die deutlichste Verbesserung, unter Verwendung des `HQS`-Konformeren-Ensembles lässt sich für die Zielstruktur P38 beobachten: Die *Sensitivität* nimmt von 0.28 (`struct_HTS`) zu 0.45 (`struct_HQS`) deutlich zu, wohingegen sich die *Spezifität* von 0.81 (`struct_HTS`) zu 0.78 (`struct_HQS`) nur minimal verschlechtert. Dieser Umstand mag zunächst verwundern, da die verwendete Referenz nicht repräsentativ für die Gesamtheit der aktiven Moleküle ist (vgl. Abschn. 18.2.3).

Ein Erklärungsansatz kann jedoch im verwendeten Pharmakophor-Modell gefunden werden: Typ I- und Typ II-Inhibitoren besitzen ein gemeinsames 'Sub'-Pharmakophor, das dem Pharmakophor der Typ I-Inhibitoren entspricht (vgl. Abb. 18.2b)²⁹⁵. Das für die Datenbanksuche verwendete Pharmakophor-Modell besitzt genau dieses Sub-Pharmakophor, das von Typ I- und Typ II-Inhibitoren erfüllt werden kann.

Bei niedrig aufgelösten Konformer-Ensembles kommt es vermutlich häufiger zu einer Kollision mit den verbotenen Sphären (die *Sensitivität* sinkt) als dies bei den hoch aufgelösten Konformer-Ensembles der Fall ist. Zusammenfassend lassen sich die Beobachtungen von KIRCHMAIR et al. bzgl. einer adäquaten Konformer-Auflösung für ROCS¹⁹⁸ auch auf konventionelle Harte-Sphären-Pharmakophore übertragen: Auch bei einer niedrigen Konformer-Auflösung lassen sich gute Ergebnisse erzielen, wenn qualitativ hochwertige Programme zur Konformerengeneration verwendet werden. Auch KIRCHMAIR et al.¹⁹⁸ verwendeten, wie auch in dieser Arbeit geschehen, OMEGA¹⁹⁹ zum Berechnen der Konformer-Ensembles.

Eine tabellarische Zusammenfassung der in diesem Abschnitt diskutierten Ergebnisse findet sich im Anhang in Abschnitt 28.2.

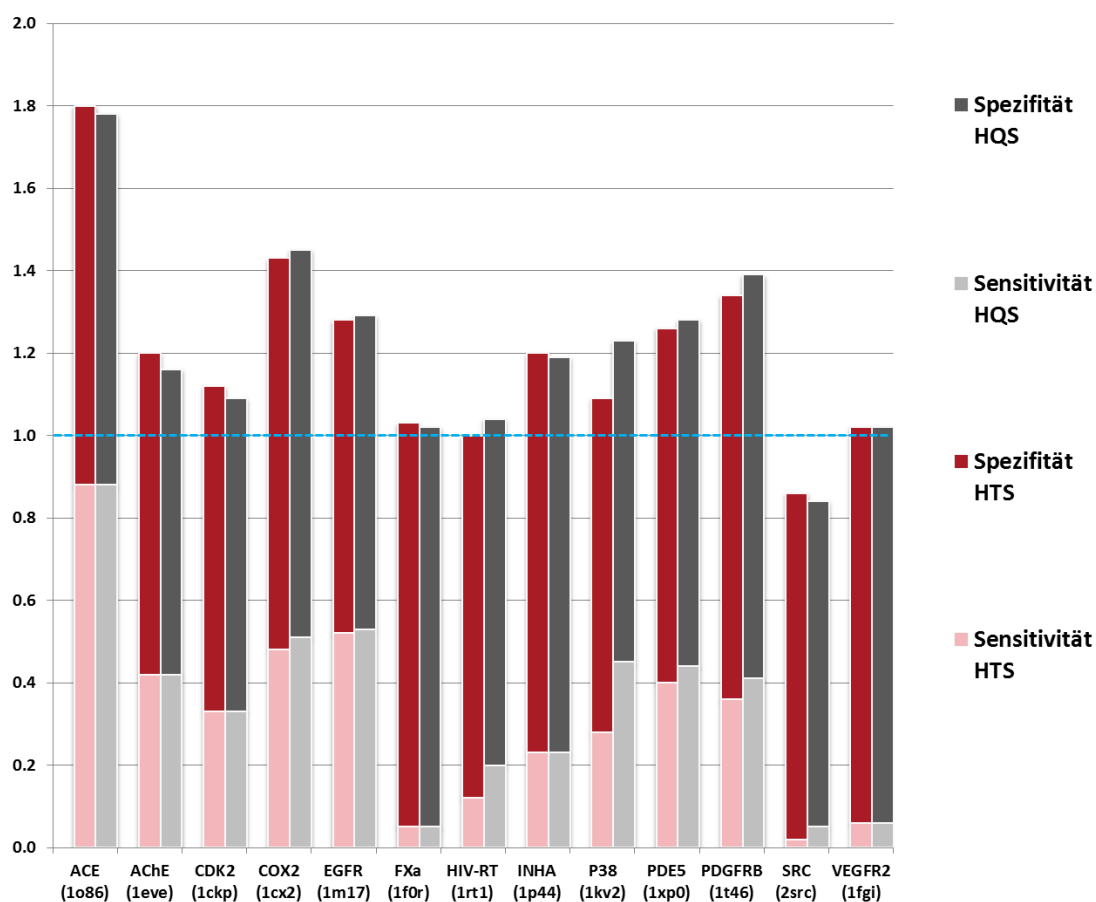


Abbildung 18.16: *Sensitivität* und *Spezifität* der manuell optimierten Pharmakophor-Modelle von CavKA oder Ligandscout für niedrig aufgelöste (struct_HTS: rot), bzw. hoch aufgelöste Konformer-Ensembles (struct_HQS: grau). Blau gepunktete Linie: Zufallsniveau.

Kapitel 19

Ergebnisse und Diskussion: MIF-basierte E-Pharmakophore

In diesem Kapitel werden die Screening-Ergebnisse der in Abschnitt 13.2 besprochenen MIF-basierten E-Pharmakophor-Methode erörtert. Als Validier-Datensatz wurde der nach Abschnitt 14.4.2.1 vorbehandelte Fieldscreen-Datensatz verwendet (`struct_HTS`). Als Screening-Referenzen wurden die nach Abschnitt 14.5.1 vorbehandelten Ligand-Rezeptor-Komplexe verwendet. Als Programm zur Datenbanksuche diente MOE. Das Diagramm in Abbildung 19.1 zeigt die Werte für *Sensitivität* und *Spezifität* für die 13 Fieldscreen-Zielstrukturen.

In ihren Analysen erzielten SALAM et al. für Pharmakophor-Modelle, die mit dem E-Pharmakophor-Ansatz abgeleitet wurden, gegenüber Modellen, die mit dem C-Pharmakophor-Ansatz nach geometrischen Kriterien erstellt worden sind, ein deutlich besseres Anreicherungsverhalten aktiver Moleküle (vgl. Abschn. 6.2.1)¹⁶⁴. Für die Zielstrukturen AChE, COX2, EGFR, HIV-RT und P38 war ein besseres Anreicherungsverhalten der manuell in MOE rekonstruierten E-Pharmakophore (blaue Säule) jedoch nicht zu reproduzieren.

Für den Datensatz HIV-RT ist mit keinem Pharmakophor-Modell, das mit CavKA, Ligandscout oder über den E-Pharmakophor-Ansatz erstellt worden ist, eine Anreicherung über Zufallsniveau erreicht worden. So drängt sich auch bei dieser Zielstruktur der Verdacht auf, dass die verwendete Referenz nicht repräsentativ für die Gesamtheit der aktiven Moleküle ist, ähnlich wie im Falle von P38 (s. Abschn. 18.2.3). Auch auf dem P38-Datensatz wurden nur Anreicherungen marginal über Zufallsniveau erzielt. Die auf den Datensätzen HIV-RT und P38 erzielten Anreicherungen sind vermutlich eher auf das Datensatzdesign zurückzuführen als auf die verwendeten Methoden.

Klar gegen die Qualität der E-Pharmakophor-Methode sprechen die auf den Zielstrukturen AChE und COX2 erzielten Ergebnisse. Auf dem Datensatz AChE werden Anreicherungen unter, auf dem Datensatz COX2 Anreicherungen auf Zufallsniveau erzielt. In beiden Fällen schneidet das MIF-basierte E-Pharmakophor (rote Säule) besser ab.

Einzig und allein auf dem EGFR-Datensatz reicht die Leistungsfähigkeit des E-Pharmakophors an die des mit CavKA erstellten Modells heran (graue Säule).

Zusammenfassend lässt sich festhalten, dass der E-Pharmakophor-Ansatz (rote Säulen) den CavKA-, bzw. Ligandscout-Modellen (graue Säulen) keineswegs überlegen ist. Die guten Ergebnisse, die von SALAM et al. erzielt worden sind, sind vermutlich auf den verwendeten Validier-Datensatz zurückzuführen: Für die Datensätze der Aktiven wurden nur wenigen Strukturen gewählt (zwischen 20 u. 33 Molekülen)¹⁶⁴. Über die Diversität der aktiven Moleküle untereinander im jeweiligen Datensatz geben die Autoren keinerlei Auskunft. Ist diese zu gering, werden scheinbar gute Ergebnisse erzielt (vgl. den von GOOD und OPREA beschriebenen *analogue bias*²⁶⁵). Des Weiteren wurden für *alle* Zielstrukturen die 1000 gleichen Decoy-Moleküle des *Glide Decoy Sets* verwendet³⁰⁶. Bei einem derartigen Vorgehen ist nicht davon auszugehen, dass die aktiven Moleküle optimal in Köder-Moleküle eingebettet sind (vgl. ROHRER und BAUMANN in Referenz¹¹). So ist es sehr wahrscheinlich, dass die von SALAM et al. beobachteten Anreicherungen letztlich artifizieller Natur sind und nur das sub-optimale Design des Datensatzes abbilden. Auch die MIF-basierten E-Pharmakophore weisen eine deutlich geringere Leistungsfähigkeit auf als die mit CavKA oder Ligandscout erstellten Modelle. Problematisch bei der automatisierten, energiebasierten Ableitung von Pharmakophor-Modellen ist, dass sowohl Scoring-Funktionen als auch insbesondere MIFs viele Positionen als 'wichtig' definieren, obgleich sie es nicht sind (*falsch-positive* Information). Ferner sind Scoring-Funktionen und MIFs in den meisten Fällen nicht in der Lage eine Position exakt vorherzusagen, sondern vielmehr mit einer gewissen Unschärfe. Eine solche Zuordnung ohne Einfluss des Anwenders durchzuführen gelingt nur bedingt, weshalb der E-Pharmakophor- und der MIF-basierte E-Pharmakophor-Ansatz ein deutlich schlechteres Anreicherungsvermögen aufweisen.

Die Anwendungsbeispiele in Abschnitt 18.1 zeigen, dass eine manuelle Zuordnung von konturierten MIFs zu korrespondierenden Eigenschaften nach visueller Inspektion in den meisten Fällen deutlich besser gelingt als eine automatisierte. Eine tabellarische Zusammenfassung der in diesem Abschnitt besprochenen Ergebnisse sowie die Angabe mit welcher Methode die *Best*-Modelle erstellt worden sind, findet sich im Anhang in Abschnitt 28.3.

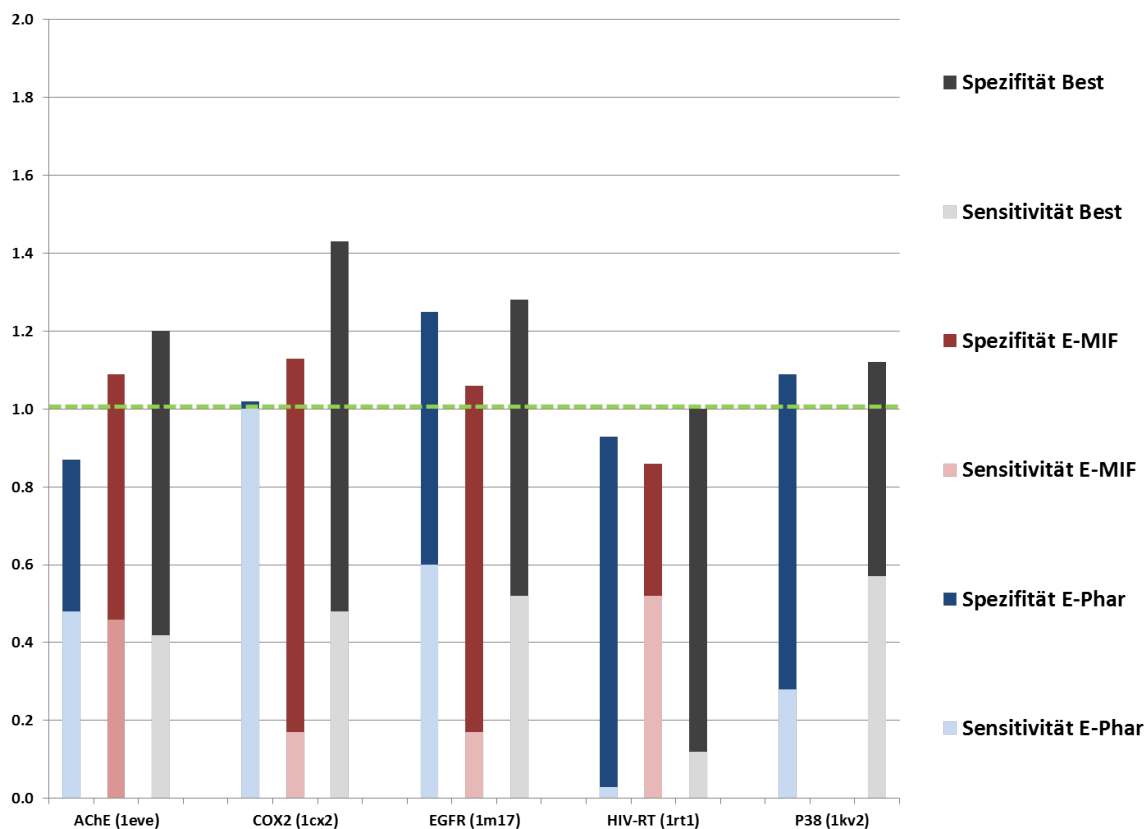


Abbildung 19.1: *Sensitivität* und *Spezifität* der Pharmakophor-Modelle des E-Pharmakophor-Ansatzes (blau) und der MIF-basierten E-Pharmakophor-Modelle (rot). Im Vergleich dazu steht das leistungsfähigste Pharmakophor-Modell (CavKA oder Ligandscout; grau). Grün gepunktete Linie: Zufallsniveau.

Kapitel 20

Ergebnisse und Diskussion: PLIF Driven Parallel Screening (PLIDriPaS)

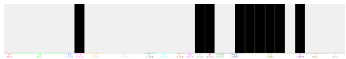









In den Abschnitten 20.1 und 20.2 werden die Ergebnisse der MOE⁶² *Protein Ligand Interaction Fingerprint* (Abk. PLIF) Berechnungen für die *Phosphodiesterase-5* (Abk. PDE5) präsentiert. Hierbei wird gezeigt, wie PLIFS genutzt werden können, um unterschiedliche Bindungsmodi zu detektieren.

In Abschnitt 20.3 werden die Screening-Ergebnisse der erstellten Modelle einzeln und nach Hitlisten-Vereinigung diskutiert. Als Validier-Datensatz wurde der nach Abschnitt 14.4.2.1 vorbehandelte Fieldscreen-Datensatz verwendet (struct_HQS). Als Screening-Referenzen wurden die nach Abschnitt 14.5.1 vorbehandelten Ligand-Rezeptor-Komplexe verwendet. Als Programm für Datenbanksuchen diente MOE.

20.1 PLIF-Berechnungen und alternative Bindungsmodi der PDE5

PLIFs wurden von unterschiedlichen PDE5-Ligand-Rezeptor-Komplexen berechnet. Die Ergebnisse der PLIF-Berechnungen sind in Tabelle 20.1 in der Spalte 'PLIF' zu finden. Eine tabellarische Zusammenfassung der in diesem Kapitel diskutierten sowie weitere Ergebnisse, die mit unterschiedlichen Pharmakophor-Modellen erzielt worden sind, sind im Anhang in Kapitel 29 aufgeführt.

Tabelle 20.1: Zehn PDE5 Ligand-Rezeptor-Komplexe aus der sc-PDB sowie *Sensitivität* und *Spezifität* für die spezifischsten, manuell optimierten Modelle. Aus jedem Cluster wurde jeweils nur ein Mitglied für das Ableiten der Pharmakophor-Modelle verwendet. *Sens.: Sensitivität; Spez.: Spezifität. Weitere Erläuterungen s. Text.*

Cluster	PDB	Name	Sens.	Spez.	PLIF
(I)	1uho	Vardenafil	-	-	
(I)	1xp0	Vardenafil	0.18	0.95	
(I)	1tbf	Sildenafil	-	-	
(I)	2h42	Sildenafil	-	-	
(I)	3jwq	Sildenafil	-	-	
(I)	3bjc	5-ethoxy-4-(1-methyl-7-oxo-3-propyl-6,7-dihydro-1H-pyrazolo[4,3-d]pyrimidin-5-yl)thiophene-2-sulfonamide	-	-	
(II)	1t9s	Guanosin-5'-Monophosphat	0.00	1.00	
(III)	1xoz	Tadalafil	0.32	0.98	
(IV)	2h44	Icarisid II	0.05	0.98	
(V)	3jwr	3-Isobutyl-1-Methylxanthin	0.32	0.87	
PLIDriPaS			0.64	0.83	

Für die PLIFs wurden insgesamt fünf Cluster erhalten: ein relativ großer Cluster (I) mit sechs sehr ähnlichen und sogar redundanten Strukturen sowie vier Cluster mit nur je einem Mitglied (sog. Singletons), die sich strukturell und anhand ihrer PLIFs klar von den Mitgliedern aus Cluster (I) unterschieden. Interessanterweise sind die PLIFs der redundanten Strukturen keineswegs identisch (1uho; 1xp0; Vardenafil, bzw. 1tbf; 2h42; 3jwq; Sildenafil), was zeigt, dass es zur Ausbildung unterschiedlicher Interaktionen kommt, die zu unterschiedlichen PLIFs führen (s. PLIFs in Tab. 20.1 für PDB 1tbf; PDB 2h42 und PDB 3jwq).

Das Zustandekommen, bzw. Nicht-Zustandekommen von Wechselwirkungen kann durch Proteinflexibilität erklärt werden. Hydrophobe Wechselwirkungen sind hierbei generell schwer vorherzusagen und weniger eindeutig zu definieren als gerichtete Interaktionen. Ferner können auch polare Gruppen, wie der 4-Methylpiperazin-Substituent in Sildenafil, erhebliche Flexibilität besitzen (vgl. Abb. 20.2). Unterschiedliche Bindungsmodi für Sildenafil (für die Ligand-Rezeptor-Komplexe PDB 1tbf und PDB 1udt) und Vardenafil (für die Ligand-Rezeptor-Komplexe PDB 1xp0 und PDB 1uho) wurden bereits von TAMINAU et al. diskutiert¹⁸⁷.

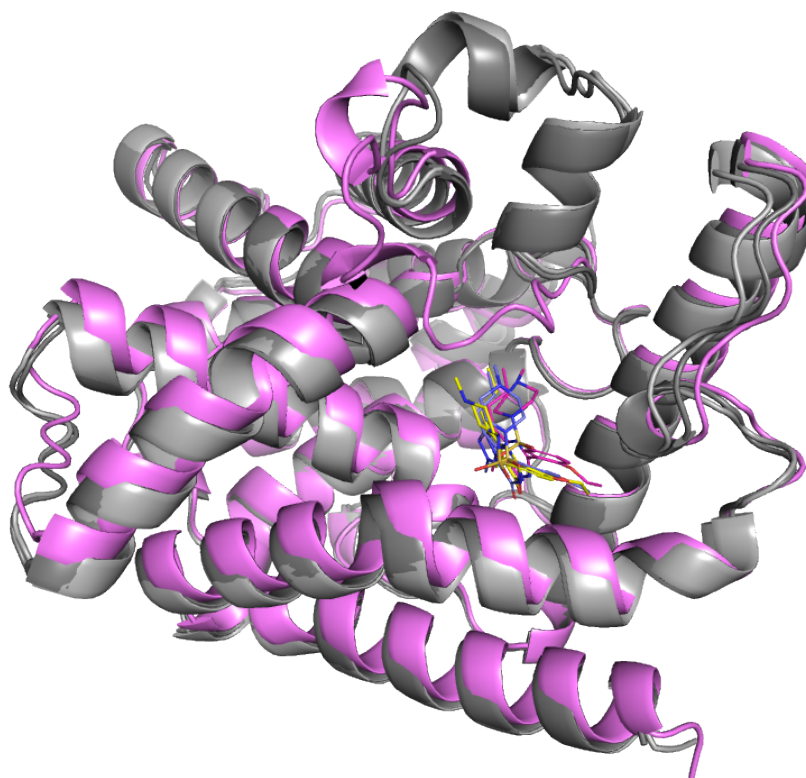


Abbildung 20.1: Drei superpositionierte Sildenafil-Ligand-Rezeptor-Komplexe. *PDB 1tbf: hellgrau; PDB 2h42: pink; PDB 3jwr: dunkelgrau.*

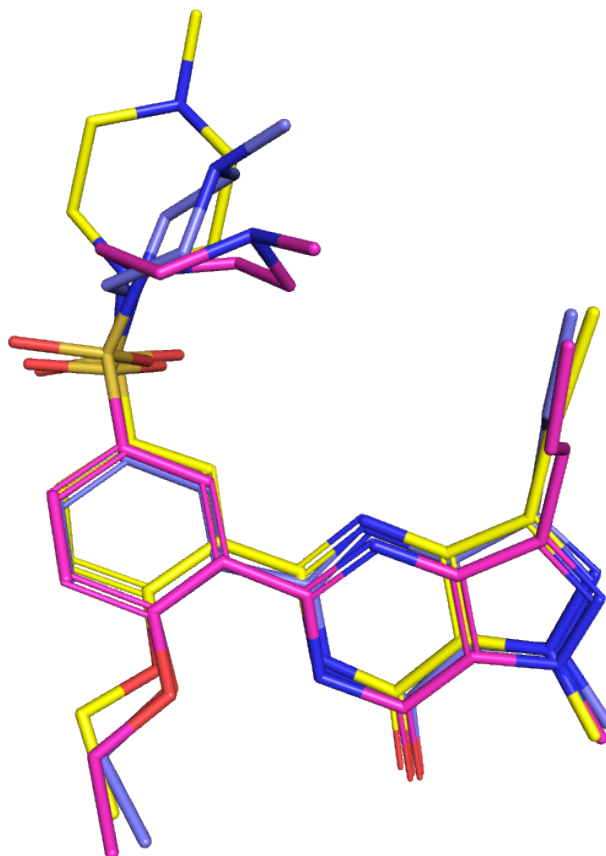


Abbildung 20.2: Orientierung der Sildenafil-Moleküle in der Bindetasche nach Superpositionierung der Ligand-Rezeptor-Komplexe. *1tbf*: blau; *2h42*: pink; *3jwr*: gelb.

Abbildung 20.1 zeigt die Überlagerung der im Rahmen dieser Arbeit verwendeten Sildenafil-Ligand-Rezeptor-Komplexe *1tbf*, *2h42* und *3jwr*. Die unterschiedlichen Orientierungen der Liganden in der Bindetasche sind in Abbildung 20.2 dargestellt. Die Tatsache, dass derselbe Ligand in der Lage ist unterschiedliche Ligand-Rezeptor-Interaktionen auszubilden, verdeutlicht, dass bei der Ableitung von Pharmakophor-Modellen auf Basis von Ligand-Rezeptor-Komplexen derselben Zielstruktur auch unterschiedliche Pharmakophor-Modelle erhalten werden können. Im folgen Abschnitt werden die Vorteile von PLIFs zur Analyse unterschiedlicher Bindungsmodi, bzw. deren Nutzen für die Auswahl von Screening Referenzen näher erläutert.

20.2 Auswahl der Referenzen für das '*Parallel Screening*'

KINNINGS und JACKSON wiesen auf den Einfluss der verwendeten Referenz auf die Screening-Ergebnisse unter Verwendung ihrer ligandzentrischen *LIGMATCH*-Methode auf dem Fieldscreen-Datensatz hin²⁶¹. In einem *Data-Mining* suchten die Autoren zunächst nach alternativen Ligand-Rezeptor-Komplexen derselben Zielstruktur und isolierten die Liganden aus ihrer Bindetasche. Der Validier-Datensatz wurde mit allen alternativen Referenzen durchsucht. Nun wurden für jeden dieser Referenz-Liganden sowie für die Aktiven und Köder-Moleküle *Openbabel Fingerprints*^{216,217} berechnet. Der Fingerprint der Referenz wurde mit den Fingerprints aller Aktiven sowie Köder-Molekülen aus dem jeweiligen Datensatz verglichen. Nach der *MAX-Fusion Rule*³⁰⁷ wurden für die finale Bewertung von Aktiven und Köder-Molekülen die *LIGMATCH*-Scores aus jener Hitliste verwendet, die den höchsten *Tanimoto-Koeffizienten* zu einer der verwendeten Referenz-Moleküle aufwiesen²⁶¹.

KIRCHMAIR et al. nutzten für ein Screening mit ROCS auf dem DUD-Datensatz anstatt eines einzigen Referenz-Moleküls eine Auswahl aktiver Moleküle mit unterschiedlichen Grundgerüsten, die als Referenz verwendet wurden, um die Ergebnisse der Datenbanksuchen zu verbessern¹⁹¹.

Beide Methoden haben gemein, dass man sich bei der Auswahl der Screening-Referenzen entweder direkt aus der Menge der aktiven Substanzen bedient (vgl. KIRCHMAIR¹⁹¹) oder ein Abgleich mit den aktiven Molekülen stattfindet (vgl. KINNINGS und JACKSON²⁶¹). In einem prospektiven Szenario ist hingegen nichts, bzw. oft nur wenig über die aktiven Moleküle bekannt.

Bei KIRCHMAIR et al.¹⁹¹ wird auf der einen Seite die Menge an aktiven Molekülen, die gefunden werden soll, reduziert. Auf der anderen Seite wird das Spektrum der Information über die aktiven Moleküle stark erweitert. Bei einer solchen Vorgehensweise sollte ein verbessertes Anreicherungsverhalten nicht verwundern.

Wird wie bei KINNINGS und JACKSON ein *Abgleich* der Referenz mit den Aktiven, bzw. den Köder-Molekülen durchgeführt²⁶¹, ist nicht ersichtlich, ob aufgrund des Fieldscreen-Datensatzdesigns ein *DUD-2D-Bias* (vgl. Abschn. 14.3.1) in die Analyse eingebracht wird, was zu einer scheinbaren Anreicherung führen kann. Um dies zu analysieren, müsste geprüft werden, ob die *Tanimoto-Koeffizienten* der 2D-*Openbabel Fingerprints* zwischen Referenz- und Datenbank-Molekül mit dem Score der 3D-Ähnlichkeit (*ROCS*/ *LIGMATCH*) korrelieren.

Ist dies der Fall, wäre von einer scheinbaren Anreicherung aktiver Moleküle auszugehen. Der in dieser Arbeit verwendete Ansatz, gezielt Ligand-Rezeptor-Komplexe über die unterschiedlichen PLIFs auszuwählen, bei denen Liganden in diversen Bindungsmodi binden, ist frei von den oben beschriebenen Problemen und bildet ein prospektives Szenario deutlich besser nach: Die aktiven Moleküle werden nicht inspiziert, auch nicht zur Auswahl oder gar als Referenz selbst verwendet. Alternativ hätten auch für alle zehn Ligand-Rezeptor-Komplexe (s. Tab. 20.1) diverse Pharmakophor-Modelle erstellt und für ein *Parallel-Screening* verwendet werden können. Dies ist bewusst vermieden worden, um die Menge an Screening-Referenzen (und damit die Informationsgrundlage) gering zu halten.

20.3 Ergebnisse PLIDriPaS (PDE5)

Beim Betrachten der Screening-Ergebnisse (s. Tab. 20.1; s. Spalte *Sens.*, bzw. *Spez.*) fällt auf, dass zwei Modelle eine sehr hohe *Sensitivität* aufwiesen: die aus den PDB-Komplexen 1xp0 (Ligand: Vardenafil) und 2h44 (Ligand: Icarisid II) erstellten Modelle waren nur in der Lage, einen sehr kleinen Teil der aktiven Moleküle anzureichern. Interessanterweise ist 1xp0 die für den Datensatz vorgeschlagene Referenz und scheint somit ebenfalls nur begrenzt repräsentativ für die Menge an aktiven Molekülen zu sein.

Das auf Basis von 1t9s abgeleitete Modell war so spezifisch, dass es nicht von einem einzigen aktiven Molekül erfüllt worden ist. Dies verwundert nicht weiter, da es sich beim co-kristallisierten Liganden um *Guanosin-5-Monophosphat* (Abk. GMP) handelt, das bei der Spaltung des nativen PDE5 Substrats, dem *cyclischen Guanosin-5-Monophosphat* (Abk. cGMP), entsteht. Bei den aktiven Molekülen des Fieldscreen-Datensatzes handelt es sich jedoch ausschließlich um Inhibitoren.

Bei den Beispielen 1xoz und 3jwr ist eine wesentlich bessere Anreicherung zu erkennen. Da die Pharmakophor-Modelle auf der Basis von Kristallstrukturen erstellt worden sind, in denen alternative Bindungsmodi vorhanden sind, enthält jede der fünf Hitlisten strukturell unterschiedliche Liganden. Daher wird nach Vereinigung der Hitlisten die erzielte Anreicherung erheblich verbessert (*Sensitivität* = 0.64; *Spezifität* = 0.83), wohingegen sich die *falsch-positiv* Rate nur wenig verschlechtert. Sicherlich ließen sich die guten Ergebnisse auch annähernd durch die Vereinigung der Hitlisten mit den auf Basis von 1xoz und 3jwr erstellten Modellen erzielen.

Ohne Inspektion der aktiven Moleküle ist bei einer Validierung jedoch nicht bekannt *wie* repräsentativ eine Referenz für den Satz an aktiven Molekülen ist. Die sehr spezifischen Modelle, die auf Basis von 1xp0 und 2h44 erstellt worden sind, tragen kaum zur Anreicherung aktiver Moleküle mit bei. Auf der anderen Seite erhöhen sie die *falsch-positiv* Rate nur unmerklich.

Eine Auflistung der Pharmakophor-Modelle, die für die Pharmakophor-Suchen in diesem Abschnitt verwendet wurden sowie weitere Pharmakophor-Modelle, die im Rahmen der durchgeführten Analysen erstellt worden sind, finden sich im Anhang in Kapitel 29.

Kapitel 21

Ergebnisse und Diskussion: CavKA^{HYBRID}

In den folgenden Abschnitten werden die Screening-Ergebnisse der in Abschnitt 13.3 besprochenen CavKA^{HYBRID}-Methode erörtert. Als Validier-Datensatz wurde der nach Abschnitt 14.4.2.1 vorbehandelte Fieldscreen-Datensatz verwendet (struct_HTS). Als Screening-Referenzen wurden die nach Abschnitt 14.5.1 vorbehandelten Ligand-Rezeptor-Komplexe verwendet. Als Programm zur Datenbanksuche diente Pharao. ROC-Kurven für die 13 Zielstrukturen, die mit den jeweiligen Pharmakophor-Modellen erhalten worden sind, befinden sich im Anhang in Abschnitt 30.1.

21.1 Ergebnisse: Vergleich von CavKA^{HYBRID} mit der ligandzentrischen Pharao-Methode

Abbildung 21.1 zeigt die $ROC_{EF1\%}$ -Werte für die 13 Fieldscreen Zielstrukturen, die mit der CavKA^{HYBRID}-Methode erzielt worden sind (vgl. Abschn. 13.3 und Abschn. 16.3). Graphen für $ROC_{EF5\%}$, bzw. $ROC_{EF10\%}$ finden sich im Anhang in Abschnitt 30.2.

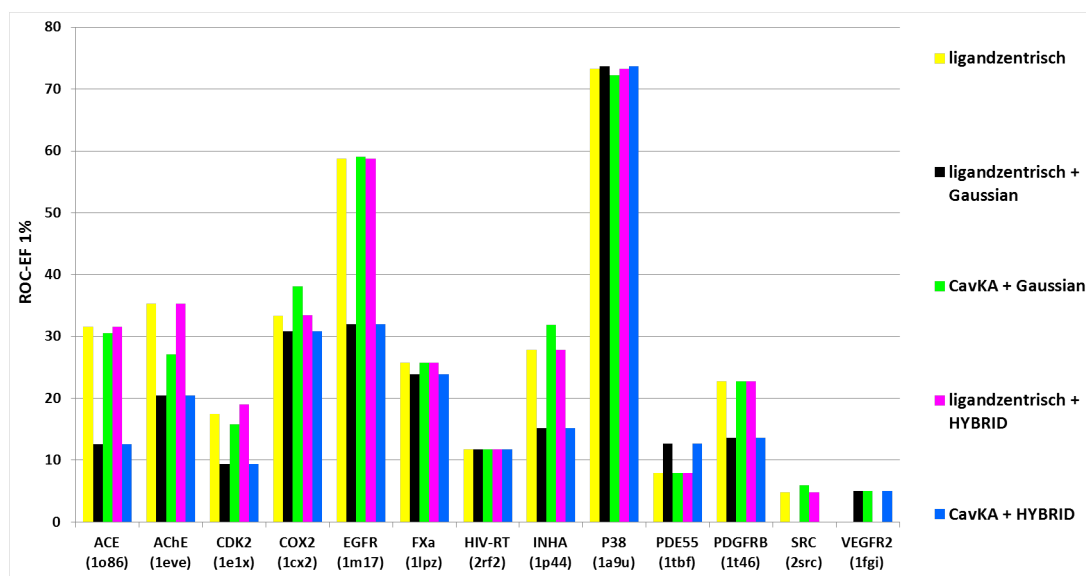


Abbildung 21.1: $ROC_{EF1\%}$ -Werte für die 13 Fieldscreen-Zielstrukturen.

Für den ausschließlich ligandzentrischen Pharaon-Ansatz (Abk. *ligandzentrisch*; gelber Balken) und den ligandzentrischen Pharaon-Ansatz, bei dem die superpositionierten Moleküle mit dem durch Harte-Sphären repräsentierten Rezeptor abgeglichen worden sind (Abk. *ligandzentrisch + HYBRID*; pinker Balken), sind nahezu identische $ROC_{EF1\%}$ -Werte zu beobachten. Selbiges ist für den strukturbasierten Ansatz zu beobachten, bei dem der Rezeptor durch verbotene Gauß-Sphären repräsentiert wird (Abk. *CavKA + Gaussian*; hellgrüner Balken), mit Ausnahme der Zielstrukturen ACE und CDK2, bei denen ein leicht schlechteres Anreicherungsverhalten festzustellen ist. Auf der Zielstruktur AChE kann ein deutlich schlechteres Anreicherungsverhalten beobachtet werden. Auf den Zielstrukturen COX2, INHA und SRC ist hingegen eine minimal höhere, auf der Zielstruktur VEGFR2 eine deutlich höhere Anreicherung zu beobachten.

Für den ligandzentrischen Ansatz, bei dem der Rezeptor durch verbotene Gauß-Sphären repräsentiert wird (Abk. *ligandzentrisch + Gaussian*; schwarzer Balken) und den CavKA^{HYBRID}-Ansatz (Abk. *CavKA HYBRID*; blauer Balken) sind in allen Fällen identische $ROC_{EF1\%}$ -Werte zu beobachten.

Verglichen mit dem ligandzentrischen Ansatz (*ligandzentrisch*) ist bei den beiden letztgenannten Methoden in nahezu allen Fällen eine deutlich geringere Anreicherung und nur in zwei Fällen (HIV-RT und FXa) eine annähernd als gleich zu bewertende Leistungsfähigkeit zu beobachten. Nur in einem Fall (PDE5) kann ein deutlich besseres Abschneiden beobachtet werden.

21.2 Diskussion: Vergleich von CavKA^{HYBRID} mit der ligandzentrischen Pharao-Methode

21.2.1 Ligandzentrische Pharao-Methode

In den meisten Fällen konnten mit der ausschließlich ligandzentrischen Pharao-Methode die besten Ergebnisse erzielt werden. In diesem Zusammenhang zeigt sich der ligandzentrische Ansatz gegenüber dem strukturbasierten Ansatz überlegen, was u.a. für die Qualität der von Pharao durchgeführten Superpositionierung der Datenbank-Moleküle spricht (vgl. dazu die Ergebnisse der ligandzentrischen Methoden in Kap. 22).

21.2.2 Verbotene Gauß-Sphären

Der massive Einbruch der Leistungsfähigkeit, verglichen mit dem ligandzentrischen Pharao Ansatz (*ligandzentrisch*), der zu beobachten ist, wenn der Rezeptor durch verbotene Gauß-Sphären repräsentiert wird (*ligandzentrisch* + *Gaussian*), lässt darauf schließen, dass diese ein hohes Maß an Rauschen in die Analyse einbringen: Ursprünglich gute Superpositionierungen werden trotz des geringen σ -Wertes der verbotenen Gauß-Sphären von 1Å (statt 1.7Å in den Pharao Standardeinstellungen) abgewertet. Bei der Betrachtung von ausschließlich interagierenden pharmakophoren Ligandeigenschaften (*CavKA* + *Gaussian*) lässt sich dieser Umstand nicht beobachten. Eventuell schlechtere Superpositionierungen (verursacht durch weniger spezifische Pharmakophor-Modelle) werden korrekt durch die verbotenen Gauß-Sphären bestraft, sodass ein Anreicherungsverhalten beobachtet werden kann, das gleichwertig, bzw. minimal schlechter als der ligandzentrische Ansatz zu bewerten ist.

21.2.3 Verbotene Harte-Sphären

Obwohl die CavKA^{HYBRID}-Pharmakophor-Modelle (*CavKA* + *Hybrid*) deutlich mehr Information enthalten als der ligandzentrische Ansatz (Information über Ligand-Rezeptor-Interaktion und die Gestalt des Rezeptors), zeigt letzterer (*ligandzentrisch*) ein deutlich besseres Anreicherungsverhalten. Da nur wenige interagierende Ligandeigenschaften bei der CavKA^{HYBRID}-Methode definiert werden, führt dies zu suboptimalen Superpositionierungen, die verglichen mit dem strukturbasierten Ansatz *CavKA* + *Gaussian* wesentlich stärker abgewertet werden, was zu einem

Einbruch der Leistungsfähigkeit führt. Werden durch Pharao superpositionierte Datenbank-Moleküle mit den verbotenen Harten-Sphären abgeglichen, kommt es zu keinen Einbußen bezüglich der Leistungsfähigkeit (*ligandzentrisch + Hybrid*). Die über Pharao superpositionierten Datenbank-Moleküle sind auf den Pharmakophor, der verglichen mit dem strukturbasierten Ansatz (*CavKA + Hybrid*) relativ viele Ligandeigenschaften besitzt, recht genau angepasst. Auch wenn pharmakophore Eigenschaften zwischen Referenz- und Datenbank-Molekül nicht optimal zur Deckung gebracht werden können, definiert der viele Eigenschaften umfassende Referenz-Pharmakophor die Orientierung der Datenbank-Moleküle in der Bindetasche, was die Gefahr von Kollisionen mit den verbotenen harten Sphären mindert. Trotz eines vermeintlichen Maßes an 'mehr' Information lassen sich mit CavKA^{HYBRID} die Screening-Ergebnisse nicht verbessern.

Kapitel 22

Ergebnisse und Diskussion: Ligandzentrische Methoden im Vergleich

In den folgenden Abschnitten werden die Screening-Ergebnisse der in Kapitel 7 besprochenen ligandzentrischen Methoden erörtert. Als Validier-Datensatz wurde der nach Abschnitt 14.4.2.2 vorbehandelte Fieldscreen-Datensatz verwendet (lig-cent_HTS). Als Screening-Referenzen wurden die nach Abschnitt 14.5.2 vorbehandelten Liganden verwendet.

22.1 Ergebnisse ligandzentrischer Methoden: Einzelne Zielstrukturen

Im Folgenden werden die einzelnen ROC-Kurven für die 13 Zielstrukturen, die unter Verwendung der jeweiligen Technik erhalten worden sind, beschrieben. Die ROC-Kurven wurden entsprechend der verwendeten Eigenschaften, die bei der Datenbanksuche berücksichtigt wurden, eingefärbt. *Shape*-basierte Ansätze sind in Magenta, Methoden, die *Elektrostatik* verwenden, in Blau und Methoden, die pharmakophore Eigenschaften anhand von Substrukturen definieren, in Cyan dargestellt. Im Folgenden werden die ROC-Kurven der einzelnen Zielstrukturen näher beschrieben.

22.1.1 ACE

Bei dem ACE-Datensatz zeigte $ROCS_{combo}$ die beste Anreicherung zu einem frühen Zeitpunkt der Datenbanksuche, $ROCS_{shape}$ schnitt im weiteren Verlauf nur wenig besser als $ROCS_{color}$ ab. Bemerkenswerterweise fielen die frei verfügbaren *shape*-basierten Methoden *Piramid* und $ShaEP_{shape}$ im Vergleich zu $ROCS_{shape}$ stark ab. Mit *Pharao* wurden bessere Ergebnisse als mit $ROCS_{color}$ und nahezu ähnlich gute Ergebnisse wie mit $ROCS_{combo}$ erzielt. EON_{combo} zeigte eine etwas bessere Leistungsfähigkeit als $ROCS_{color}$. Die schlechtesten Resultate wurden mit $ShaEP_{combo}$ erzielt. Zu beachten ist, dass eine erhebliche Anreicherung mit der molekularen Masse als Deskriptor (*Mass*) erzielt worden ist.

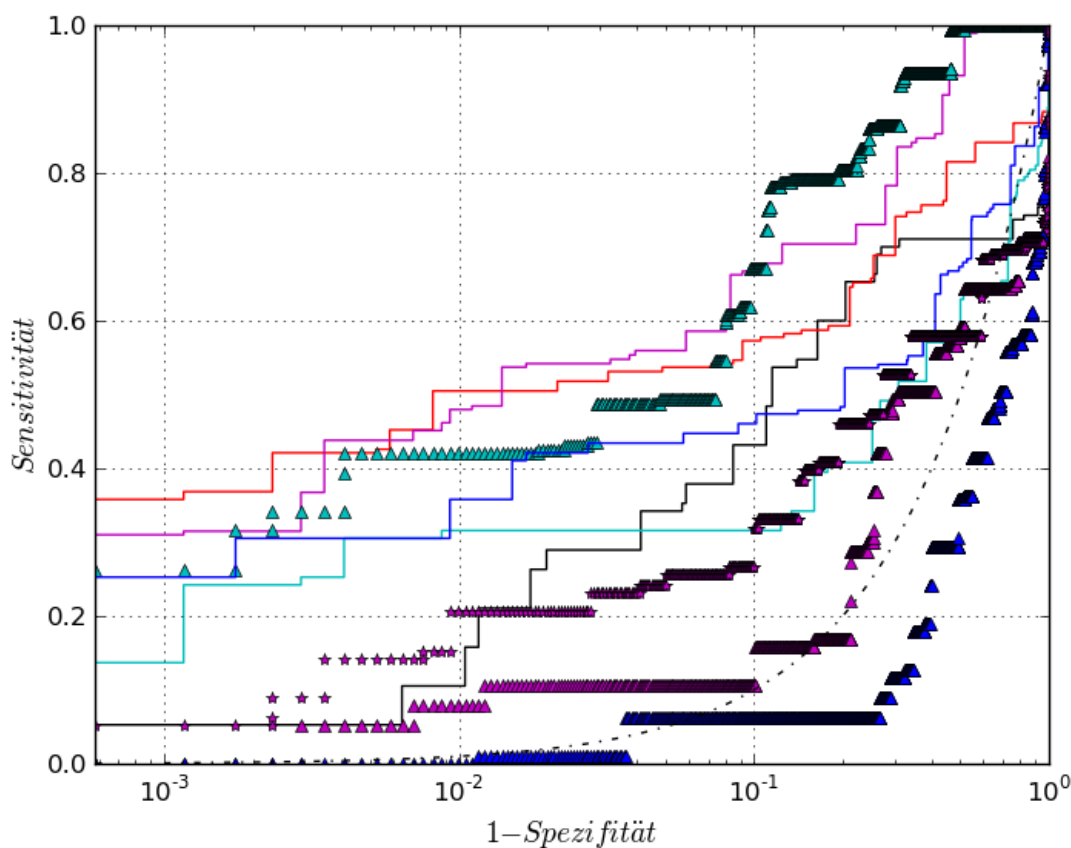


Abbildung 22.1: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für ACE-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.2 AChE

Die besten Anreicherungen auf dem AChE-Datensatz wurden mit *Pharao* erzielt, welche erheblich besser waren als die mit $ROCS_{color}$ erzielten Ergebnisse. Die beobachteten Anreicherungen von $ROCS_{combo}$ und $ROCS_{shape}$ waren sehr ähnlich. Die Leistungsfähigkeit von $ShaEP_{shape}$ war nur wenig schlechter als die von $ROCS_{shape}$, wohingegen *Piramid* als *shape*-basierter Ansatz deutlich abfiel. EON_{combo} konnte nicht an die mit $ROCS_{combo}$ erzielten Ergebnisse anknüpfen. Auch wenn EON_{combo} eine bessere Anreicherung zu einem frühen Zeitpunkt als $ShaEP_{combo}$ aufwies, schnitt $ShaEP_{combo}$ im späteren Verlauf besser ab.

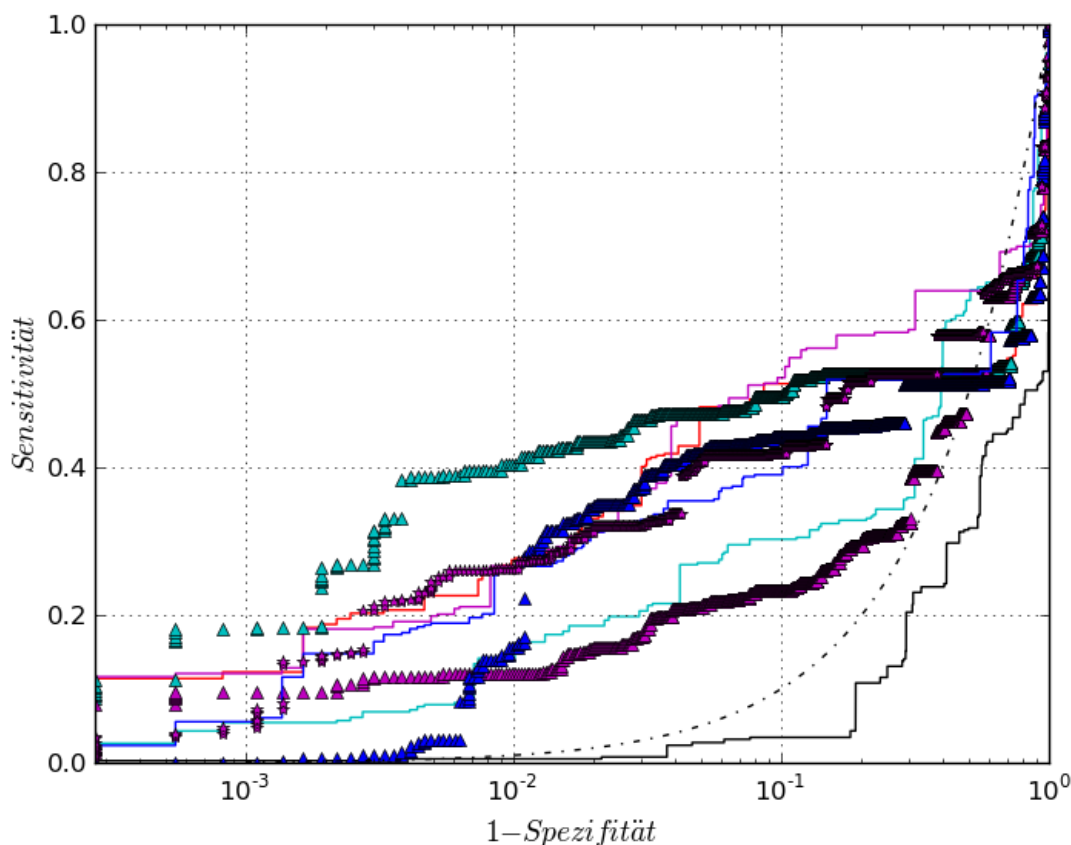


Abbildung 22.2: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für AChE-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.3 CDK2

Die Screening-Ergebnisse der betrachteten Methoden variierten auf dem CDK2-Datensatz wenig, und die ROC-Kurven wiesen einen qualitativ ähnlichen Verlauf auf. $ROCS_{combo}$ schnitt am besten ab, *Pharao* hingegen am schlechtesten. Es ist zu betonen, dass mit dem *Mass*-Deskriptor auf CDK2-Datensatz eine beträchtliche Anreicherung (besonders zu frühen Zeitpunkten der Datenbanksuche) erzielt worden ist, die über weite Strecken sogar besser als die mit $ROCS_{combo}$ erzielten Ergebnisse war.

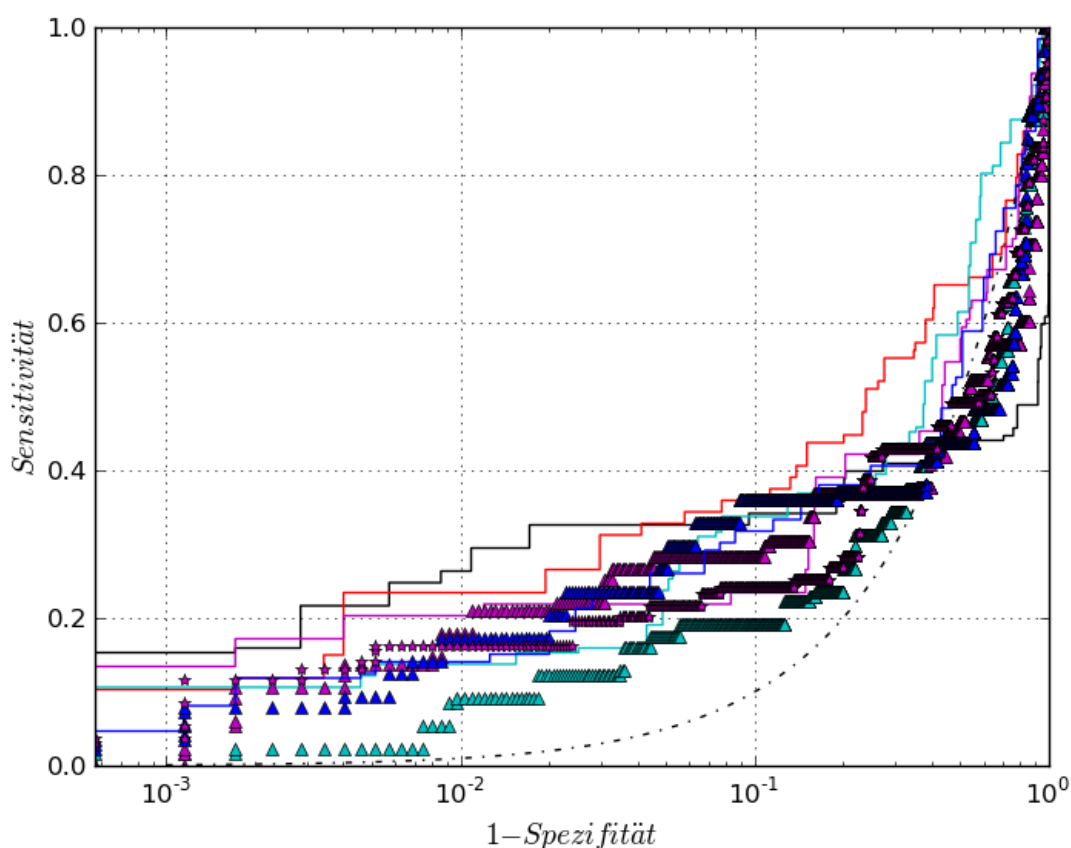


Abbildung 22.3: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für CDK2-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.4 COX2

Auch auf dem COX2-Datensatz wurden mit $ROCS_{combo}$ die besten Ergebnisse erzielt. $Pharao$, $ROCS_{color}$, $ROCS_{shape}$, $Piramid$, $ShaEP_{shape}$ und EON_{combo} erzielten eine ähnliche Anreicherung aktiver Moleküle, wohingegen die Leistungsfähigkeit von $ShaEP_{combo}$ deutlich abfiel (unter Zufallsniveau). Ferner war zu beobachten, dass die Methoden, die pharmakophore Eigenschaften regelbasiert definieren, besser abschnitten als *shape*-basierte Methoden. Letztgenannte schnitten tendenziell besser ab als Methoden, die elektrostatische Potentiale zur Definition pharmakophorer Eigenschaften verwenden.

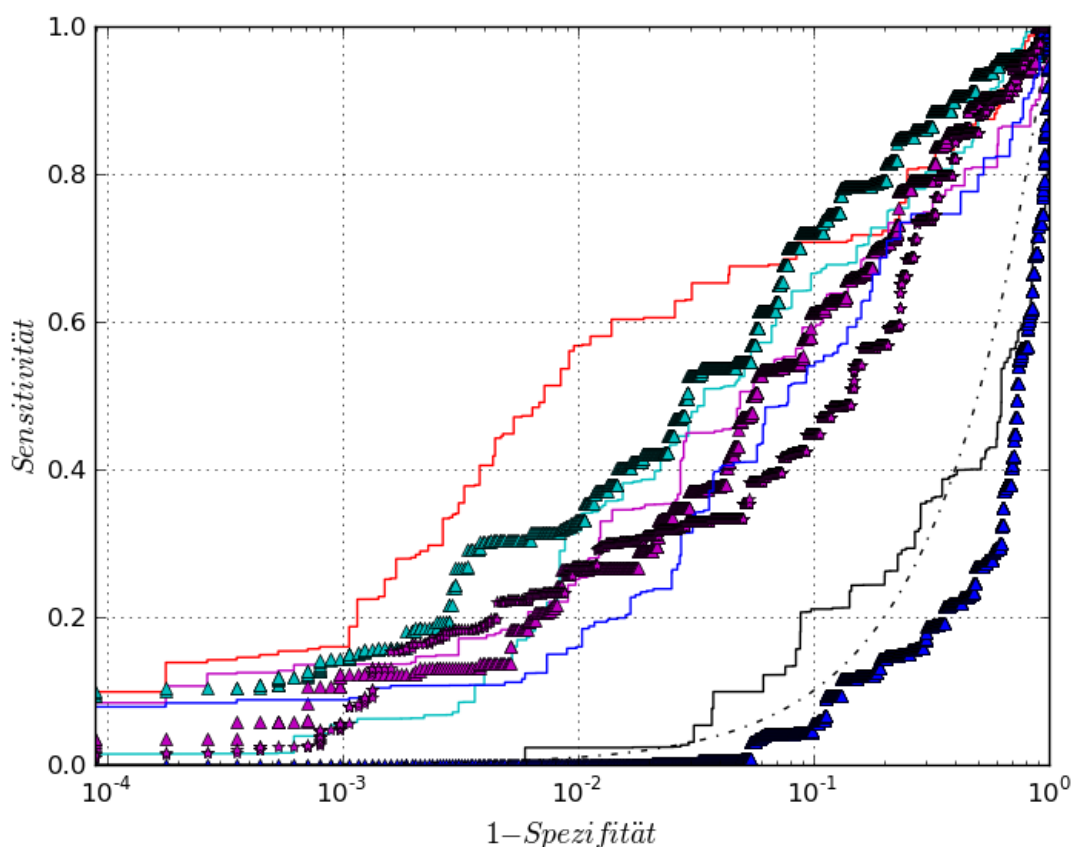


Abbildung 22.4: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für COX2-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: $Pharao$; magenta Dreieck: $Piramid$; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: $Mass$; schwarz gepunktet: Zufall.

22.1.5 EGFR

Auf dem EGFR-Datensatz zeigte *Pharao* erneut die beste Anreicherung in frühen Stadien der Datenbanksuche und demonstrierte eine ähnliche Leistungsfähigkeit wie $ROCS_{combo}$ und $ROCS_{shape}$. Bemerkenswerterweise erzielte $ROCS_{color}$ eine deutlich schlechtere Anreicherung aktiver Moleküle. Die Leistungsfähigkeit von *Piramid* und $ShaEP_{shape}$ konnte bei Weitem nicht an die mit $ROCS_{shape}$ erzielten Ergebnisse heranreichen: Die mit $ShaEP_{shape}$ erzielten Ergebnisse befanden sich unter Zufallsniveau. $ShaEP_{combo}$ erzielte mit $ROCS_{color}$ vergleichbare Ergebnisse. Die Leistungsfähigkeit von EON_{combo} fiel im Vergleich hierzu deutlich ab.

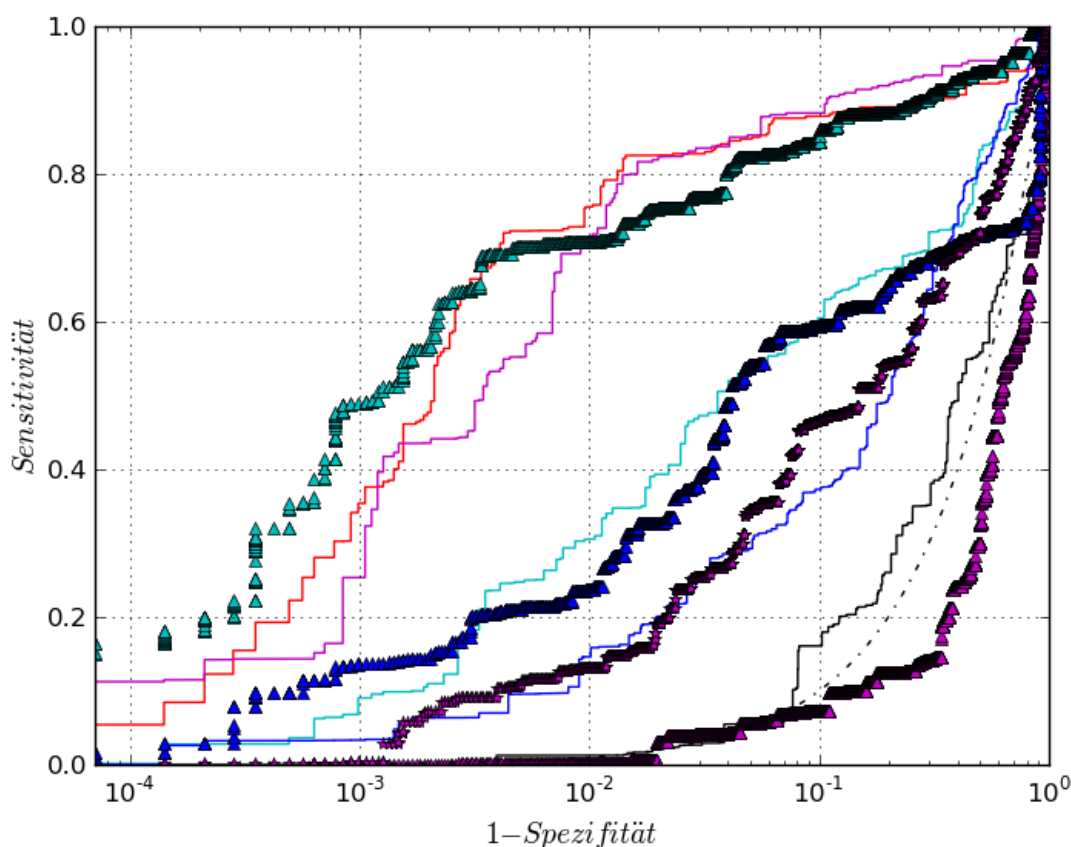


Abbildung 22.5: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für EGFR-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.6 FXa

Auf dem FXa-Datensatz war ein qualitativ ähnlicher Verlauf der ROC-Kurven für alle Methoden zu beobachten, der sich zu großen Teilen auf Zufallsniveau bewegte. Die *ShaEP_{shape}* Methode zeigte die beste Anreicherung aktiver Moleküle der verglichenen Methoden, welche jedoch auch nur marginal über dem Zufallsniveau lag.

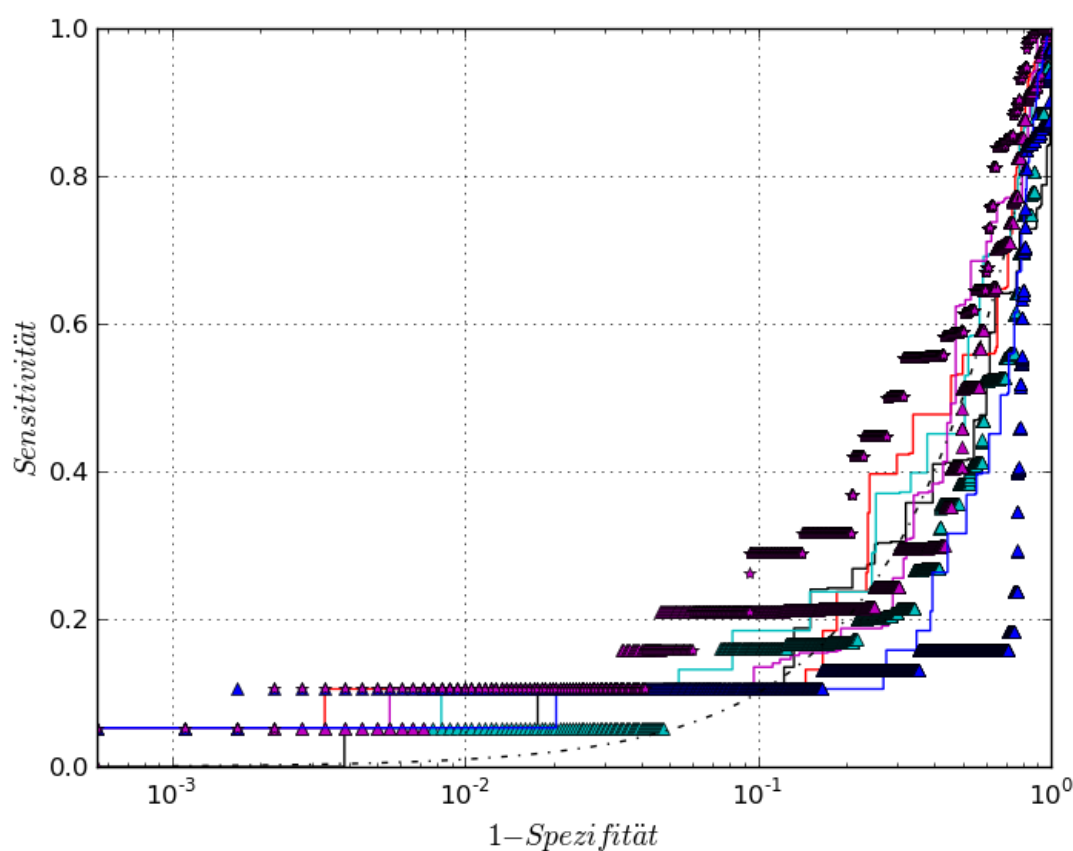


Abbildung 22.6: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für FXa-Datensatz. Rot: *ROCS_{combo}*; cyan: *ROCS_{color}*; magenta: *ROCS_{shape}*; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: *EON_{combo}*; blau Dreieck: *ShaEP_{combo}*; magenta Stern: *ShaEP_{shape}*; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.7 HIV-RT

Auch auf dem HIV-RT-Datensatz war der qualitative Verlauf der ROC-Kurven recht ähnlich. Die beste Anreicherung in frühen Stadien der Datenbanksuche wurde mit $ROCS_{shape}$ erzielt. Bis zu einem Bereich von 10% der durchsuchten Datenbank zeigten $ROCS_{shape}$, $ROCS_{combo}$, $ShaEP_{shape}$, $ShaEP_{combo}$, EON_{combo} und *Pharao* nahezu identische Ergebnisse, wohingegen die mit $ROCS_{color}$ erzielten Ergebnisse eine deutlich größere Leistungsfähigkeit aufwiesen. Die mit *Piramid* erzielten Ergebnisse (bis 10% der durchsuchten Datenbank) bewegten sich auf Zufallsniveau.

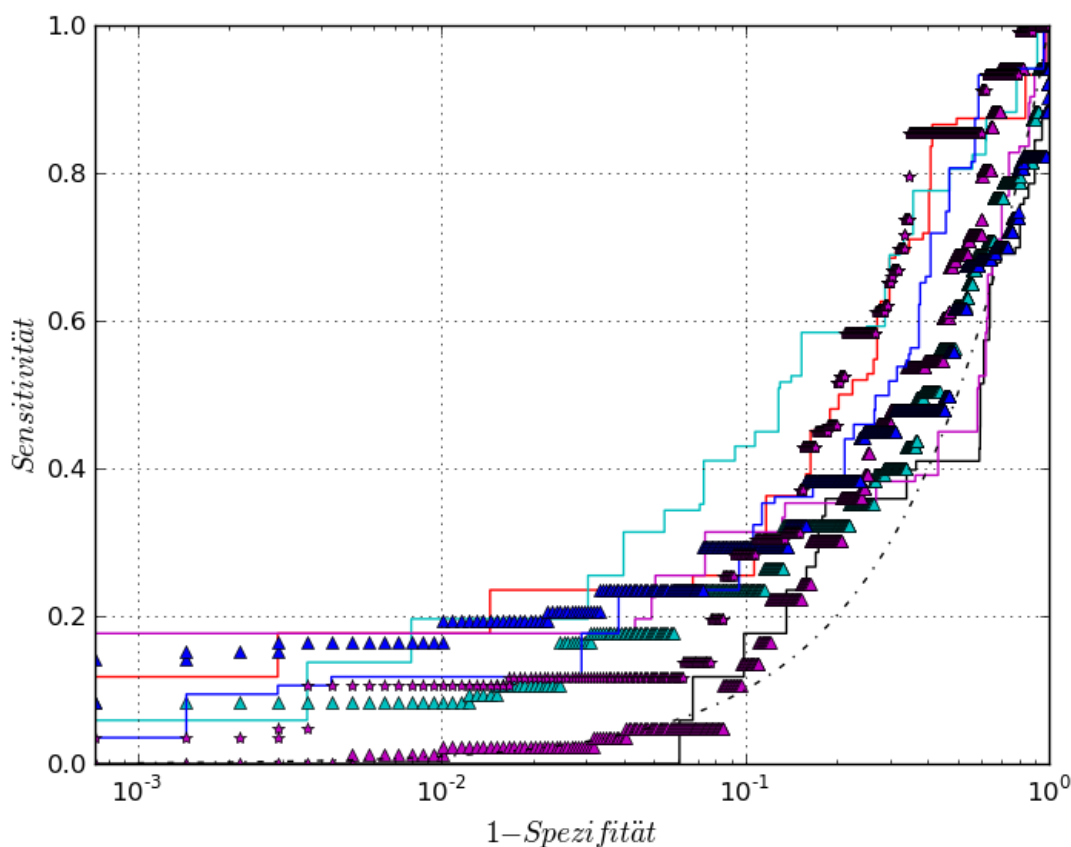


Abbildung 22.7: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für HIV-RT-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.8 INHA

Auf dem INHA-Datensatz zeigte *Pharao* mit großem Abstand, im Vergleich zu den restlichen Methoden, das beste Anreicherungsverhalten aktiver Moleküle während der gesamten Datenbanksuche. Es folgten $ROCS_{shape}$ und $ROCS_{combo}$. Die Leistungsfähigkeit von EON_{combo} , $ShaEP_{shape}$ und $ROCS_{color}$ waren vergleichbar. $ShaEP_{combo}$ zeigte bis ca. 10% der zu durchsuchenden Datenbank ein mit EON_{combo} vergleichbares Anreicherungsverhalten, das im weiteren Verlauf jedoch abfiel. Die schlechteste Methode im Vergleich war *Piramid*. Obwohl eine sehr geringe Anreicherung zu Beginn der Datenbanksuche erzielt worden ist, bewegte sich die Leistungsfähigkeit zum größten Teil auf Zufallsniveau.

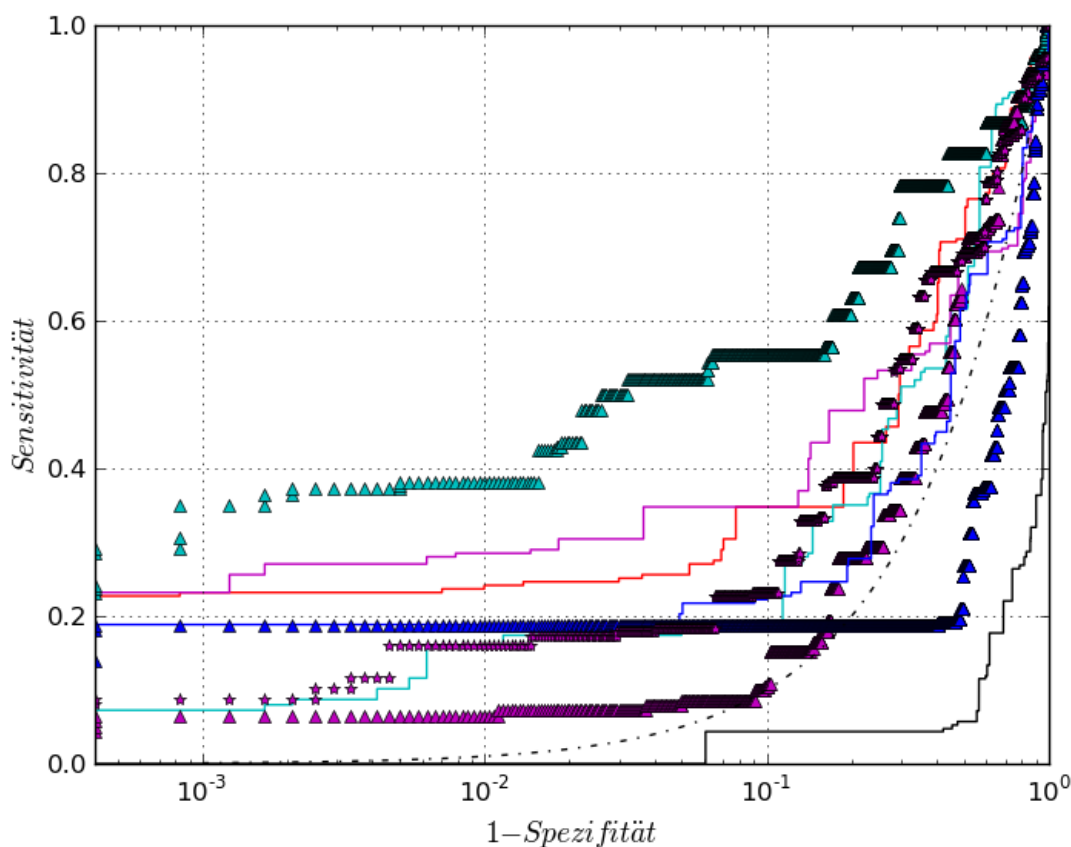


Abbildung 22.8: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für INHA-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.9 P38

Auch auf dem P38-Datensatz war ein qualitativ sehr ähnlicher Verlauf der ROC-Kurven zu beobachten, der nur schwach über dem Zufallsniveau lag. Die beste Anreicherung aktiver Moleküle wurde mit *ShaEP_{combo}* erzielt.

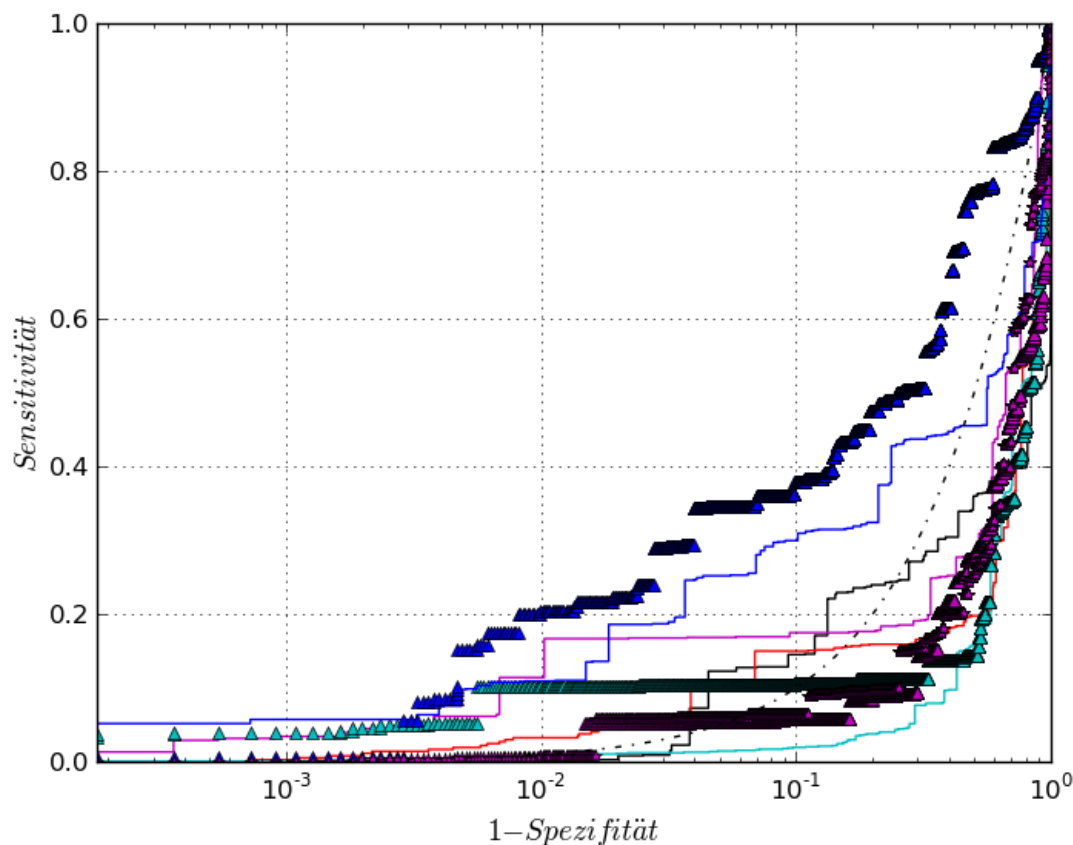


Abbildung 22.9: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für P38-Datensatz. Rot: *ROCS_{combo}*; cyan: *ROCS_{color}*; magenta: *ROCS_{shape}*; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: *EON_{combo}*; blau Dreieck: *ShaEP_{combo}*; magenta Stern: *ShaEP_{shape}*; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.10 PDE5

Auf dem PDE5-Datensatz wurde mit den meisten Methoden nur eine sehr geringe Anreicherung aktiver Moleküle erzielt, die sich größtenteils auf, bzw. unter Zufallsniveau bewegte. *Pharao* demonstrierte hingegen eine deutlich bessere Leistungsfähigkeit über das gesamte Screening. Die mit *ShaEP_{combo}* erhaltenen Ergebnisse reichten nicht an die von *Pharao* heran, befanden sich jedoch bis 10% der durchsuchten Datenbank deutlich über Zufallsniveau.

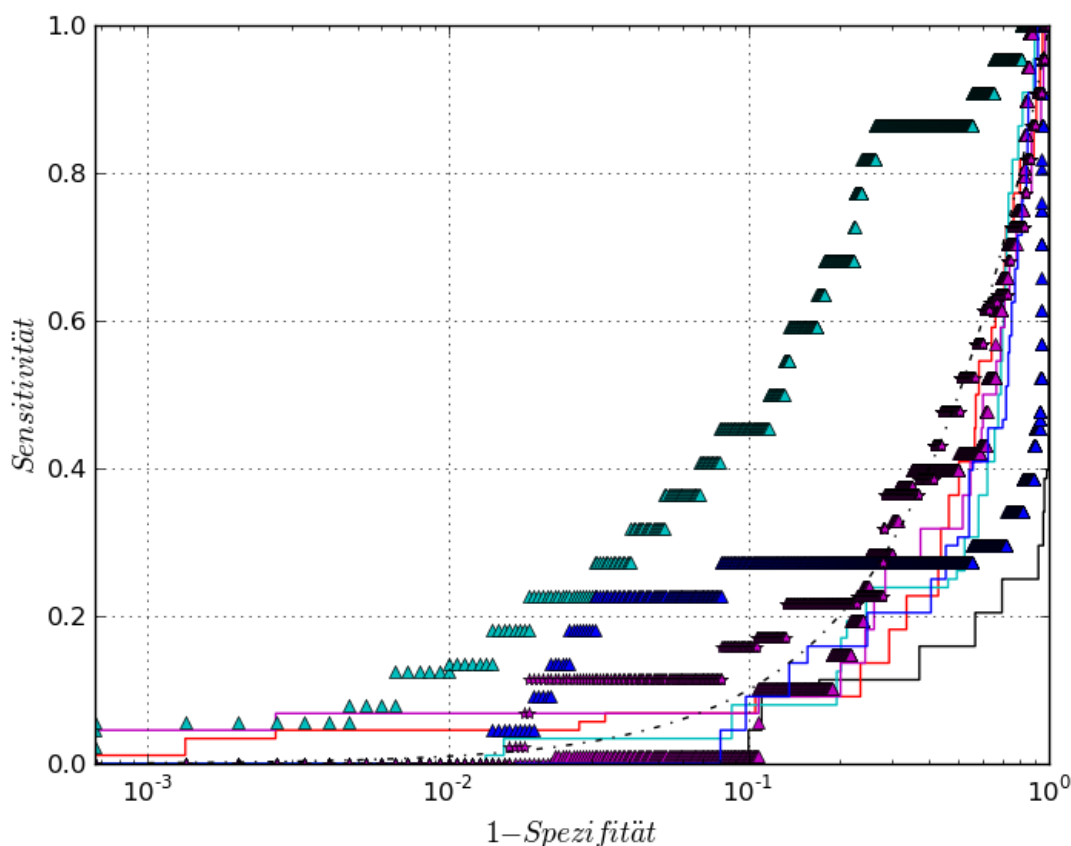


Abbildung 22.10: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für PDE5-Datensatz. Rot: *ROCS_{combo}*; cyan: *ROCS_{color}*; magenta: *ROCS_{shape}*; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: *EON_{combo}*; blau Dreieck: *ShaEP_{combo}*; magenta Stern: *ShaEP_{shape}*; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.11 PDGFRB

Auf dem PDGFRB Datensatz wurde mit EON_{combo} über weite Strecken die beste Anreicherung aktiver Moleküle erzielt. Die mit $ShaEP_{combo}$ erhaltenen Ergebnisse waren im direkten Vergleich sehr viel schlechter. Von den *shape*-basierten Methoden zeigte $ROCS_{shape}$ die beste Anreicherung in frühen Stadien der Datenbanksuche. Die mit *Piramid* und $ShaEP_{shape}$ erzielten Ergebnisse waren annähernd gleichwertig. Im weiteren Verlauf des Screenings näherten sich die ROC-Kurven der *shape*-basierten Methoden jedoch stark an. Die Ergebnisse von Pharaos waren nahezu mit denen von $ROCS_{shape}$ identisch. Die Leistung von $ROCS_{color}$ war vergleichbar mit *Piramid* und $ShaEP_{shape}$. Auch auf dem PDGFRB-Datensatz ist eine beträchtliche Anreicherung aktiver Moleküle unter Verwendung der molekularen Masse als Deskriptor (*Mass*) erzielt worden.

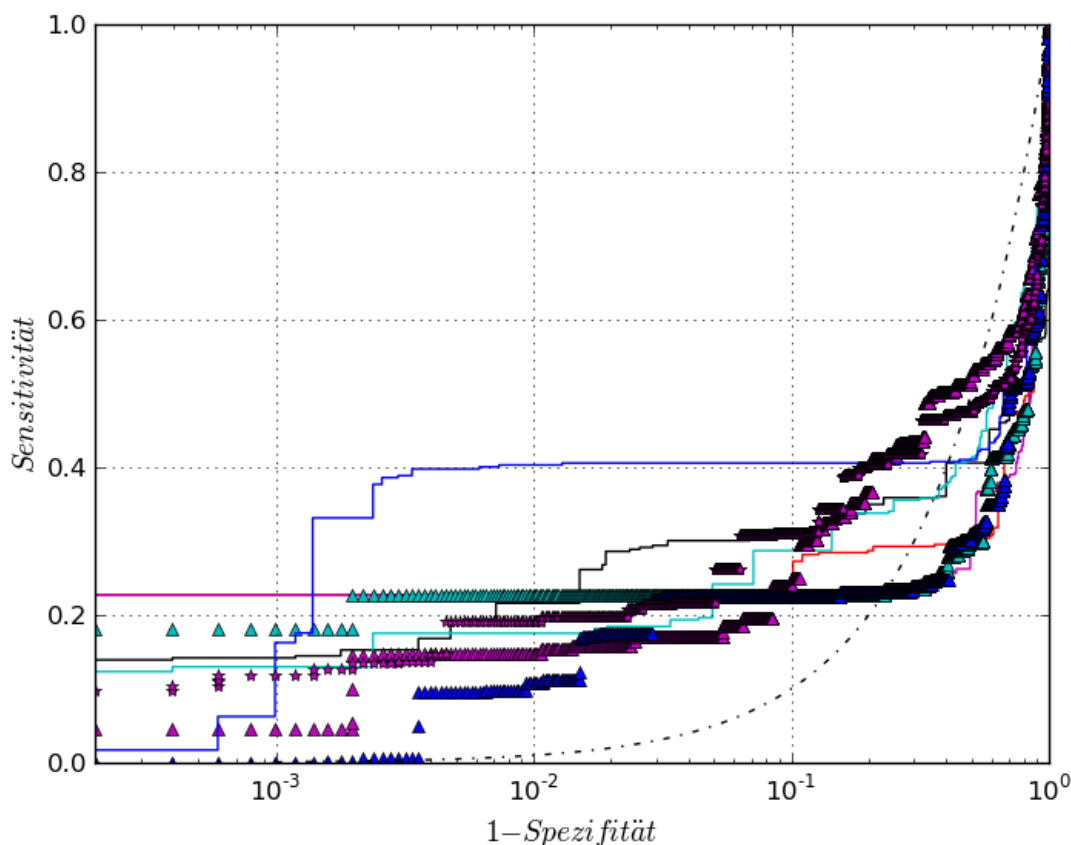


Abbildung 22.11: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für PDGFRB-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: Pharaos; magenta Dreieck: Piramid; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.1.12 SRC

Auf dem SRC-Datensatz erzielten die zu vergleichenden Methoden allenfalls eine sehr geringe Anreicherung aktiver Moleküle: $ROCS_{shape}$ und $Pharao$ zeigten eine geringe Anreicherung schwach über dem Zufallsniveau. Auch auf dem SRC-Datensatz konnte eine geringe Anreicherung unter Verwendung der molekularen Masse als Deskriptor ($Mass$) erzielt werden.

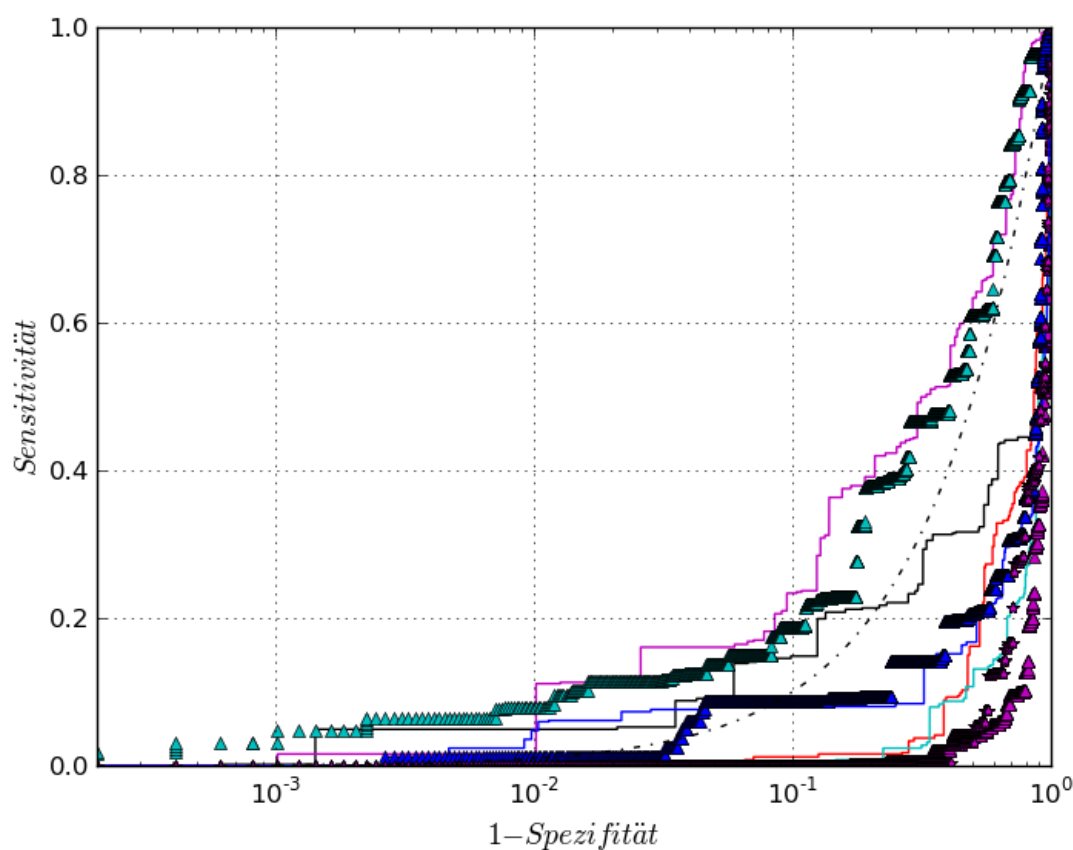


Abbildung 22.12: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für SRC-Datensatz. Rot: $ROCS_{combo}$; cyan: $ROCS_{color}$; magenta: $ROCS_{shape}$; cyan Dreieck: $Pharao$; magenta Dreieck: $Pyramid$; blau: EON_{combo} ; blau Dreieck: $ShaEP_{combo}$; magenta Stern: $ShaEP_{shape}$; schwarz: $Mass$; schwarz gepunktet: Zufall.

22.1.13 VEGFR2

Auf dem VEGFR2-Datensatz wurde die höchste Anreicherung aktiver Moleküle, die sich zudem deutlich von den anderen Ergebnissen abhebt, mit der molekularen Masse als Deskriptor (*Mass*) erzielt. *ShaEP_{shape}* war mit großem Abstand hinter der molekularen Masse die zweitbeste Methode. *Pharao* war die schlechteste Methode im Vergleich. Deren Leistungsfähigkeit lag deutlich unter Zufallsniveau. Die Leistung der restlichen Methoden bewegte sich auf oder schwach über dem Zufallsniveau.

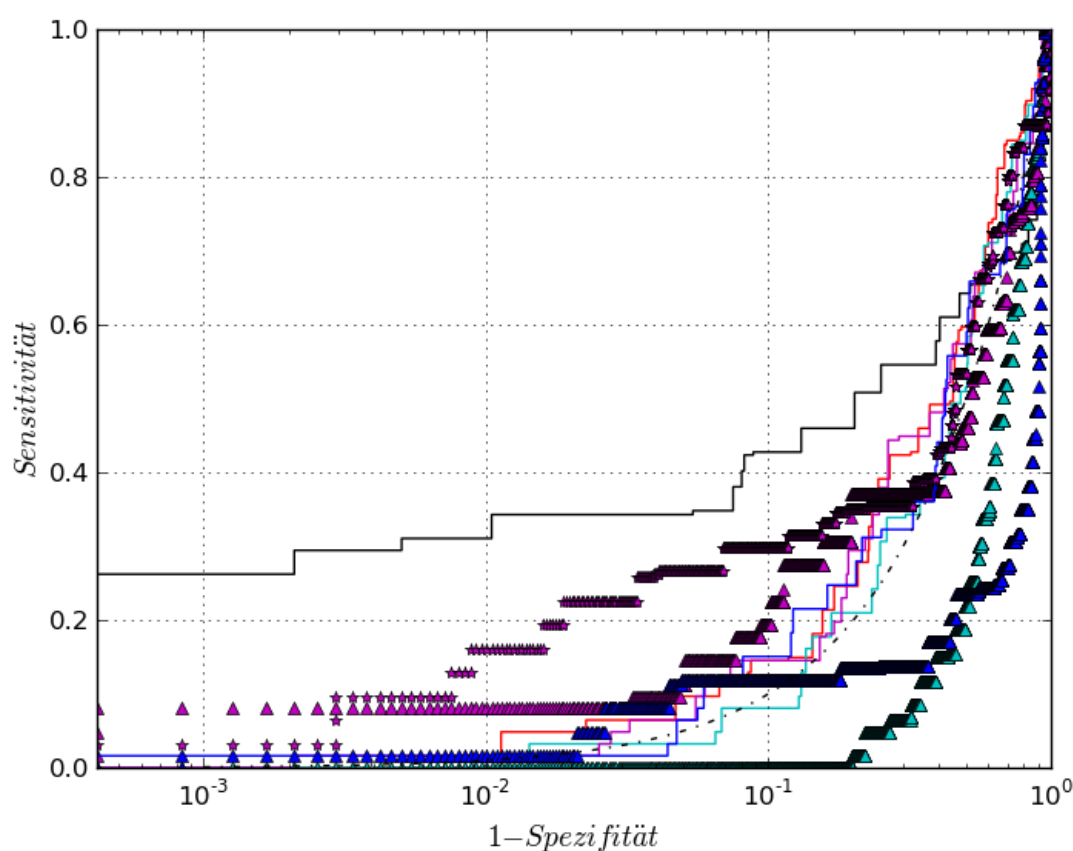


Abbildung 22.13: Screening-Ergebnisse unterschiedlicher ligandzentrischer Methoden für VEGFR2-Datensatz. Rot: *ROCS_{combo}*; cyan: *ROCS_{color}*; magenta: *ROCS_{shape}*; cyan Dreieck: *Pharao*; magenta Dreieck: *Piramid*; blau: *EON_{combo}*; blau Dreieck: *ShaEP_{combo}*; magenta Stern: *ShaEP_{shape}*; schwarz: *Mass*; schwarz gepunktet: Zufall.

22.2 Ergebnisse ligandzentrischer Methoden: Methodenvergleich

Um die ligandzentrischen Methoden objektiv miteinander vergleichen zu können, wurden sogenannte *ROC-Enrichments* (Abk. ROC-EF) für ein, fünf und zehn Prozent berechnet und die Ergebnisse in sogenannten *Heatmaps* (s. Abb. 22.14, 22.15 u. 22.16) dargestellt.

Das größte Anreicherungsverhalten aktiver Moleküle bei $ROC_{EF1\%}$ konnte bei den meisten Methoden für die Zielstrukturen ACE, COX2, EGFR und PDGFRB beobachtet werden. Bei $ROC_{5\%}$ und $ROC_{EF10\%}$ zeigten die meisten Methoden relativ gute Anreicherungen auf den Zielstrukturen ACE, COX2 und EGFR. Bei $ROC_{5\%}$ war darüber hinaus ein gutes Anreicherungsverhalten für die Zielstruktur AChE zu beobachten. Die ROC-EF-Werte nahmen generell stark im Verlauf des Screenings ab.

Von den *shape*-basierten Methoden wurden in den meisten Fällen die besten Ergebnisse mit $ROCS_{shape}$ vor $ShaEP_{shape}$ und *Piramid* erhalten. Bezüglich der Methoden, die elektrostatische Potentiale verwenden, um pharmakophore Eigenschaften zu definieren, zeigten sich EON_{combo} und $ShaEP_{combo}$ in etwa gleichwertig. Bei den Methoden, die pharmakophore Eigenschaften regelbasiert definieren, zeigte *Pharao* im Vergleich zu $ROCS_{color}$ in den meisten Fällen eine bessere Leistungsfähigkeit.

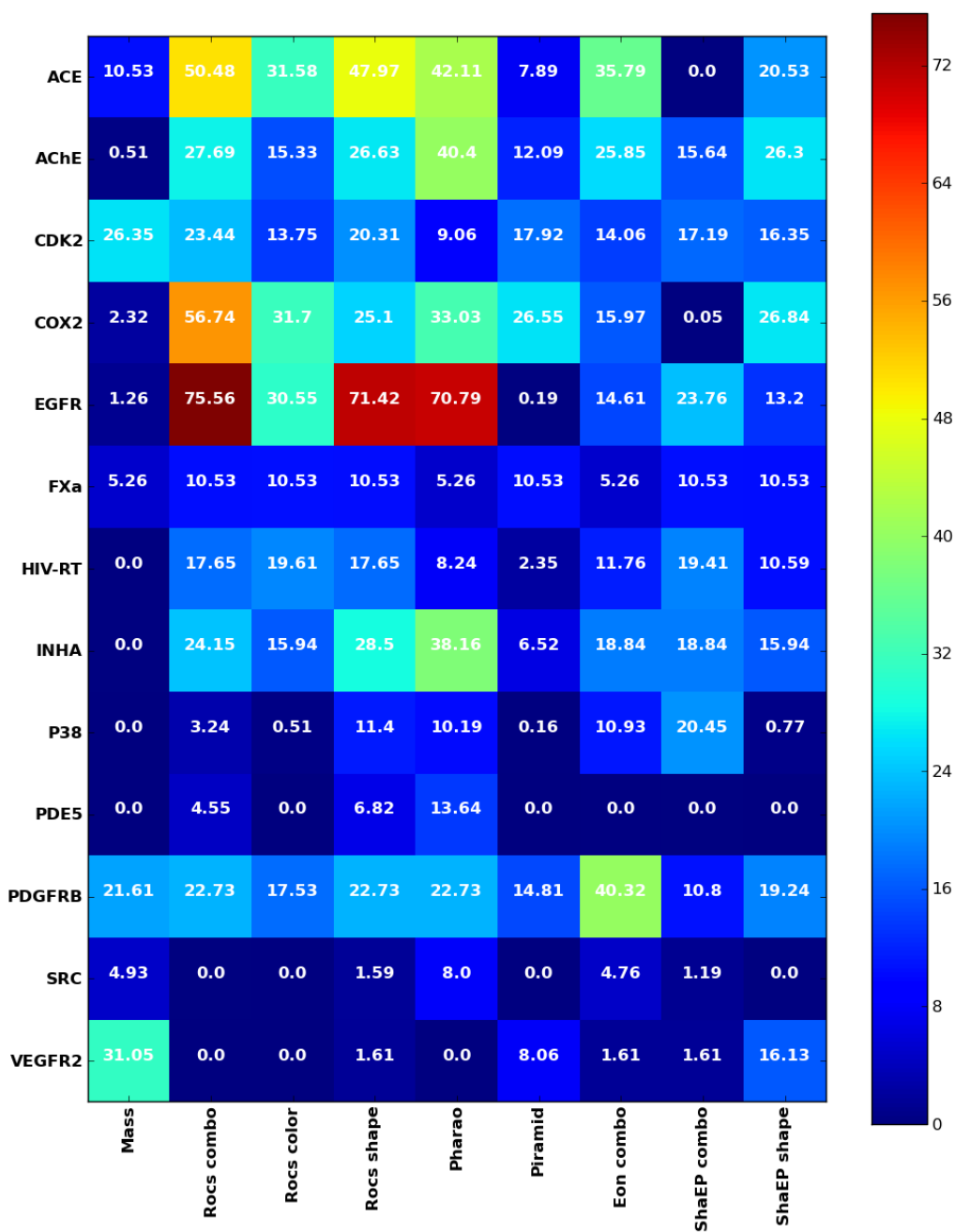


Abbildung 22.14: $ROC_{EF1\%}$ -Werte für die 13 Fieldscreen-Zielstrukturen unter Verwendung unterschiedlicher ligandzentrischer Ansätze.

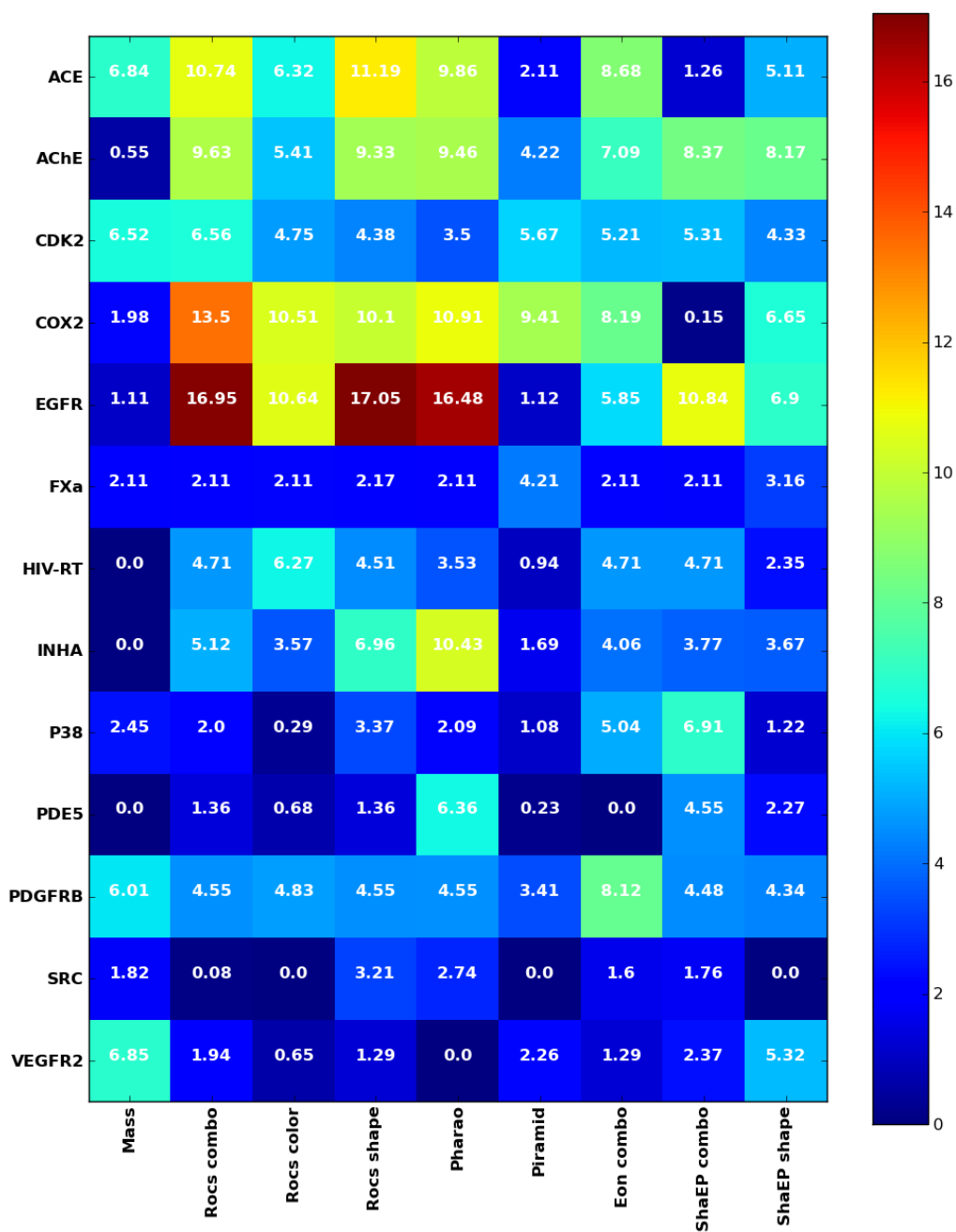


Abbildung 22.15: $ROC_{EF5\%}$ -Werte für die 13 Fieldscreen-Zielstrukturen unter Verwendung unterschiedlicher ligandzentrischer Ansätze.

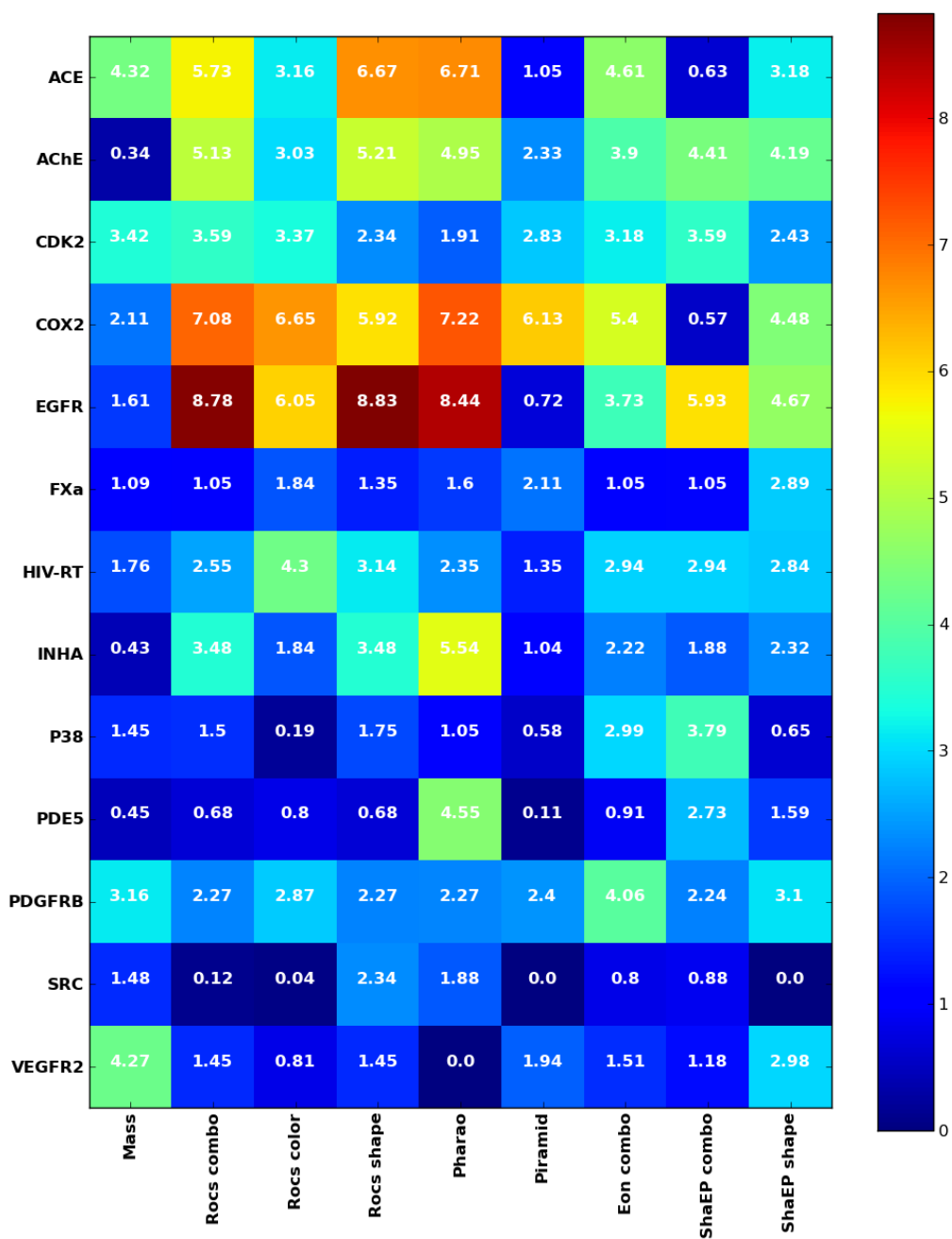


Abbildung 22.16: $ROC_{EF10\%}$ -Werte für die 13 Fieldscreen-Zielstrukturen unter Verwendung unterschiedlicher ligandzentrischer Ansätze.

22.3 Diskussion ligandzentrischer Methoden

22.3.1 Datensatzkritik und Massen Bias

Der DUD-, bzw. Fieldscreen-Datensatz (s. Abschn. 14.3.1 u. 14.3.3) ist nach Aussage der Autoren ursprünglich so konzipiert worden, dass in einem Virtuellen Screening mit einfachen physikalisch-chemischen Deskriptoren keine Anreicherungen aktiver Moleküle zu erzielen sein sollen. Somit war es umso überraschender, dass für die Datensätze ACE, CDK2, PDGFRB und VEGFR2 genau dies unter Verwendung der molekularen Masse als Deskriptor (*Mass*), welcher eigentlich nur zu Kontrollzwecken verwendet worden ist, zu beobachten war.

Der *Mass*-Deskriptor erzielte auf dem VEGFR2-Datensatz sogar höhere Anreicherungen als jede der verwendeten Screening-Methoden. Konsequenterweise sollten daher alle erzielten Ergebnisse in Relation zu denen betrachtet werden, die mit dem *Mass*-Deskriptor erzielt worden sind (vgl. BENDER und GLEN³⁹). Ferner stellt sich die Frage, wie es zu dem beobachteten Phänomen kommen konnte, da sich sowohl die DUD²⁵⁹- als auch die Fieldscreen¹⁸⁴- Autoren gezielt um einen Angleich von Aktiven und Köder-Molekülen bezüglich der einfachen physikalisch-chemischen Eigenschaften bemüht haben.

Um eine Screening Methode adäquat beurteilen zu können, empfiehlt es sich für die Datensätze, die besonders stark von diesem *Massen-Bias* betroffen sind, einen erneuten Angleich von Referenz-, Aktiven und Köder-Molekülen bzgl. der molekularen Masse und einfacher physikalisch-chemischer Eigenschaften vorzunehmen und alternative Köder-Moleküle zu verwenden.

Zusätzlich zu den bereits genannten Problemen (fehlerhafte Röntgenkristall-Strukturen als Referenz (s. Abschn. 18.1.1); fehlerhafte .sdf Dateien (s. Abschn. 14.4.1); *Analogue Bias*²⁶⁵; nicht in einem biologischen Assay bestätigte Inaktivität der Köder-Moleküle¹¹; verwendete Homologie-Modelle anstatt co-kristallisierter Ligand-Rezeptor-Komplexe¹⁸⁴) ist anzumerken, dass viele der in Fieldscreen¹⁸⁴ vorgeschlagenen Referenz-Liganden nicht repräsentativ für die aktiven Moleküle im Datensatz sind (vgl. dazu Abschn. 18.2). Auch auf diesem Datensatz konnte in vielen Fällen unter Verwendung etablierter Screening-Techniken nur eine Anreicherung auf Zufallsniveau erzielt werden. Zusammenfassend lässt sich schließen, dass der Fieldscreen-Datensatz in seiner derzeitigen Form nur in sehr begrenzter Weise sowohl als Validier-Datensatz für ligandzentrische Ansätze, als auch für strukturbasierte Ansätze verwendet werden kann.

22.3.2 Leistungsfähigkeit der ligandzentrischen Methoden

Grundsätzlich ist festzustellen, dass keine Methode auf *allen* Zielstrukturen eine überlegene Leistungsfähigkeit demonstriert. Die mit der jeweiligen Technik erzielten Ergebnisse hängen enorm von der untersuchten Zielstruktur ab.

Generell können mit Methoden, die pharmakophore Eigenschaften regelbasiert definieren (*ROCS* und *Pharao*), die besten Ergebnisse erzielt werden. Bemerkenswerterweise wird für *Pharao* vielfach ein besseres Anreicherungsverhalten als mit *ROCS_{color}* beobachtet, obwohl beiden Methoden augenscheinlich das gleiche Prinzip zugrunde liegt. Abweichungen können mit unterschiedlichen Implementierungen von Superpositionierung und regelbasierter Definitionen pharmakophorer Eigenschaften, bzw. der Direktionalität bei den *Pharao*-Pharmakophor-Modellen erklärt werden. Auch wenn die Anreicherungen (insbesondere zu sehr frühen Stadien der Datenbanksuche) von *ROCS_{combo}* und *ROCS_{shape}* im Screening Verlauf stellenweise deutlich über denen der *Pharao*-Methode lagen, so ist die Leistungsfähigkeit der frei verfügbaren *Pharao*-Methode doch beachtlich und im Großen und Ganzen vergleichbar mit den kommerziellen Methoden. Auch die Leistungsfähigkeit von *ROCS_{combo}* und *ROCS_{shape}* hängt stark von der Zielstruktur ab. In Übereinstimmung mit KIRCHMAIR et al.¹⁹¹ werden mit *ROCS_{combo}* sehr gute und robuste Ergebnisse erzielt. Der Umstand, dass die meisten Arbeitsgruppen die Kombination von molekularen Volumen (*shape*) und regelbasierten pharmakophoren Eigenschaften (*color*) in der ROCS-Scoring-Funktion für die Evaluierung ihrer Screening-Ergebnisse verwenden^{184,185,191,260,261,308}, spricht darüber hinaus für sich.

Obwohl mit ROCS superpositionierte Datenbank-Moleküle als Grundlage für die EON-Berechnungen verwendet wurden, unterscheiden sich die Ergebnisse von *ROCS_{combo}* und *EON_{combo}* erheblich. Hierbei wurden mit *ROCS_{combo}* deutlich bessere Ergebnisse erzielt. Die für EON verwendeten ROCS-Superpositionierungen wurden mit der Option '-eon_input' erzeugt, welche speziell für EON optimierte Überlagerungen der Moleküle ermöglicht. Um den tatsächlichen Einfluss der Superpositionierung zu analysieren, wurden EON-Berechnungen zum einen mit der ROCS '-eon_input'-Option und zum anderen mit *ROCS_{combo}* überlagerten Datenbankmolekülen durchgeführt. Abbildung 31.1 in Abschnitt 31.1 im Anhang zeigt, dass der Einfluss der Superpositionierung nur eine marginale Auswirkung auf die Screening-Ergebnisse hat. Somit ist das schlechtere Anreicherungsverhalten EONs weniger auf die vorausgegangene Superpositionierung zurückzuführen als vielmehr auf die Definition von Ligandeigenschaften über elektrostatische Potentiale.

Im Vergleich zu der regelbasierten *ROCS_{color}* Eigenschafts-Definition (die für schwere Atome definiert wird), ist bei Elektrostatik-Berechnungen die Definition von H-Atomen erforderlich.

Bei ROCS wird die Überlappung des Volumens, bzw. pharmakophorer Eigenschaften zwischen Referenz- und Datenbank-Molekül maximiert. Bei EON erfolgt hingegen keine Optimierung der Überlappung elektrostatischer Eigenschaften. Hieraus (erforderliche Definition der H-Atome; keine Maximierung der Überlappung) ergibt sich eine sehr viel stärkere Abhängigkeit von der verwendeten Konformeren-Auflösung. In den durchgeführten Analysen ist eine sehr niedrige Konformeren-Auflösung von maximal 100 Konformeren pro Molekül verwendet worden. Da bereits kleine Änderungen der Molekülkonformation zu sehr starken Änderungen des elektrostatischen Feldes führen können (wohingegen das Volumen von diesen minimalen Änderungen weitestgehend unbeeinflusst bleibt)¹⁸⁶, kann in einer ungenügenden Konformeren-Auflösung eine mögliche Ursache für das schlechtere Abschneiden EONs gefunden werden. Ferner ist zu erwarten, dass mit aufwändigeren Methoden berechnete Partialladungen bessere Ergebnisse liefern würden.

Auch VENKATRAMAN et al. verglichen in einem Virtuellen Screening die Leistungsfähigkeit unterschiedlicher Methoden, aktive Moleküle auf dem DUD-Datensatz anzureichern²⁶⁰. Untersucht wurden die 3D-Methoden ShaEP, EON und ROCS unter Verwendung unterschiedlicher Scoring-Funktionen sowie verschiedener 2D-Fingerprints. Hierbei ist anzumerken, dass DUD wegen des zugrundeliegenden Datensatzdesigns ungeeignet für 2D-Suchen ist (s. Abschn. 14.3.1). Darüber hinaus verwendeten die Autoren für die 3D-Methoden anstatt einer erforderlichen Multikonformeren-Datenbank eine Datenbank, die nur ein Konformer pro Molekül enthielt. Aufgrund dieser mangelhaften Datensatz-Vorbereitung lassen sich die Ergebnisse (insbesondere für EON) mit denen im Rahmen dieser Arbeit gewonnenen nicht vergleichen.

TIKKAINEN et al. verwendeten *ROCS_{color}* und *EON_{combo}* (jeweils in Standard-Einstellungen) für ihre Analysen auf dem MUV-Datensatz²⁶⁷. Obwohl die verwendeten Konformere eine deutlich höhere Auflösung (max. 200 Konformere pro Molekül; OMEGA Standard-Einstellungen) besaßen, wurden in den meisten Fällen mit *ROCS_{color}* bessere Ergebnisse erzielt.

ShaEP_{combo} und *EON_{combo}* zeigen starke Unterschiede auf unterschiedlichen Zielstrukturen in der Fähigkeit aktive Moleküle im Rahmen eines Virtuellen Screenings anzureichern. Auch wenn beide Methoden elektrostatische Potentiale in ihrer Bewertungsfunktion nutzen, so ist die zugrundeliegende Methodik doch sehr unterschiedlich:

Während EON durch ROCS überlagerte Moleküle verwendet und ein elektrostatisches Potential für das gesamte Molekül berechnet, berechnet ShaEP ein elektrostatisches Potential nur an einigen geometrisch definierten Stellen (sogenannten 'Extrempunkten')¹⁸⁵.

Daher ist bei ShaEP von einer noch größeren Abhängigkeit der Screening-Ergebnisse bezüglich der Ausgangskonformation auszugehen. Ferner werden bei ShaEP im Rahmen der Optimierung der Überlappung der elektrostatischen Eigenschaften (s. Abschn. 7.5) die verwendeten Gewichte nicht negativ bewertet. Dies hat zur Folge, dass nur eine Ähnlichkeit der Eigenschaften bewertet wird, eine Unähnlichkeit der Eigenschaften hingegen nicht zu einer Abwertung führt¹⁸⁵.

Zudem werden in der ShaEP-Methode erheblich mehr Parameter definiert. So werden numerische Grenzwerte für den lokalen Umgebungsdeskriptor und für das elektrostatische Potential festgelegt, die bestimmen, wann zwei zu vergleichende Knoten des Feldgraphen als übereinstimmend bewertet werden. Fehlerhafte Paarungen können in der sich anschließenden Optimierungsprozedur nicht kompensiert werden. Aufgrund der genannten methodischen Unterschiede verwundern die unterschiedlichen Ergebnisse daher nicht. Auch bei *ShaEP_combo* wäre eine Verbesserung der Screening-Ergebnisse zu erwarten, wenn die Konformer-Auflösung erhöht werden würde.

In der Original-ShaEP-Publikation verglichen die Autoren ihre Methode unter anderem gegen *ROCS_combo* unter Verwendung des Fieldscreen-Datensatzes. Hierbei wurden in sieben von 13 Fällen die besseren Ergebnisse mit *ShaEP_combo* erzielt. Jene Ergebnisse lassen sich jedoch mit denen in dieser Arbeit gewonnenen nicht vergleichen, da die ShaEP-Autoren auf die Verwendung des *chemotype enrichments*²⁷¹ verzichteten. Dies führt dazu, dass der im Datensatz inhärente *analogue bias*²⁶⁵ abgebildet wird. Darüber hinaus wurden andere Programme zur Datensatz-Vorbehandlung verwendet.

22.3.3 Abhängigkeit der Screening-Ergebnisse von der Referenz

Wie bereits im Rahmen der Pharmakophor-Suchen diskutiert, besteht auch bei ligandzentrischen Ansätzen eine starke Abhängigkeit der Screening-Ergebnisse von der verwendeten Screening-Referenz. Bei den Ligand-Rezeptor-Komplexen, die zur Pharmakophor-Ableitung und der sich anschließenden Datenbanksuche genutzt wurden, war für viele Datensätze nur eine geringe Anreicherung aktiver Moleküle zu beobachten (FXa, CDK2, HIVRT, P38, SRC und VEGFR2). Auch die ligandzentrischen Ansätze zeigen auf diesen Zielstrukturen ein eher mäßiges Anreicherungsverhalten.

Bei Pharmakophor-Modellen, die auf Basis von Ligand-Rezeptor-Komplexen erstellt werden, gibt der Ligand den Spielraum für mögliche Ligand-Rezeptor-Interaktionen vor und somit auch den Bereich, in dem pharmakophore Eigenschaften platziert werden können. So lassen sich die tendenziell ähnlichen Ergebnisse für die einzelnen Zielstrukturen, die mit Pharmakophor-Suchen und ligandzentrischen Ansätzen erhalten worden sind, auf die 'Ligandabhängigkeit' der beiden Methoden zurückführen.

Auch hierbei gilt: Ist die verwendete Referenz nicht repräsentativ für die Gesamtheit der aktiven Moleküle (wie z.B. FXa oder P38), werden aktive Moleküle entweder gar nicht oder nur sehr begrenzt angereichert. Somit ist die Verwendung alternativer Screening-Referenzen für die oben genannten Zielstrukturen unabdingbar.

Alternativ kann auf Screening-Referenzen, wie im Rahmen der PLIDriPaS-Methode in Abschnitt 20.2 beschrieben, bereits publizierter Studien zurückgegriffen werden. Eine weitere Möglichkeit besteht darin, alternative co-kristallisierte Liganden aus anderen Ligand-Rezeptor-Komplexen derselben Zielstruktur zu verwenden. Die sc-PDB²³¹ ist aufgrund der zugrundeliegenden Qualitätskriterien eine gute Ressource hierfür. Auch bei ligandzentrischen Ansätzen kann die PLIDriPaS-Methode verwendet werden: Werden Datenbanksuchen mit unterschiedlichen Liganden durchgeführt, die in alternativen Bindungsmodi an ihre Zielstruktur binden, ist mit einer Verbesserung der Screening-Ergebnisse nach der Vereinigung der unterschiedlichen Hitlisten zu rechnen.

Kapitel 23

Zusammenfassung

23.1 Zusammenfassung CavKA

Mit CavKA gelang die Implementierung einer Methode, die basierend auf Ligand-Rezeptor-Komplexen Pharmakophor-Modelle ableitet. Zur Modell-Erstellung wurden geometrische Kriterien (s. Kap. 11) sowie konturierte Molekulare Interaktionsfelder (MIFs) der GRID-Methode (s. Kap. 12) als eine energetische Komponente verwendet. Es konnte gezeigt werden, dass die mit diesem kombinierten Ansatz erstellten Modelle wichtige Schlüssel-Interaktionen zwischen Ligand und Bindetasche in ihrem Modell abbilden. CavKA-Pharmakophor-Modelle zeigten vielfach ein besseres Anreicherungsverhalten von aktiven Molekülen als Ligandscout-Pharmakophor-Modelle, die nach rein geometrischen Kriterien erstellt werden (s. Kap. 18).

MIFs in Pharmakophor-Modellen zu berücksichtigen, bietet gegenüber rein geometrischen Ansätzen entscheidende Vorteile bei der Identifikation relevanter hydrophober Interaktionen (s. Abschn. 18.1.1). Des Weiteren sind MIFs eine wertvolle Hilfe bei der Identifikation relevanter Wassermoleküle (vgl. Abschn. 18.1.2) und können darüber hinaus genutzt werden, um eine Priorisierung von interagierenden Ligandeigenschaften vorzunehmen. Der zuletzt genannte Punkt besitzt eine besondere Relevanz, wenn viele Ligandeigenschaften mit der Bindetasche interagieren (vgl. Abschn. 18.1.3).

Die erzielten Screening-Ergebnisse wurden enorm von der verwendeten Referenz beeinflusst. Schlechte Screening-Ergebnisse lassen sich häufig dadurch erklären, dass die verwendete Referenz nicht repräsentativ für die Gesamtheit der aktiven Moleküle ist (vgl. Abschn. 18.2).

Die Auflösung der Konformer-Ensembles beeinflusste die Ergebnisse der Datenbanksuchen nur wenig. Bereits mit einer niedrigen Auflösung sind gute Ergebnisse erzielt worden (s. Abschn. 18.3), was klar für die Qualität des verwendeten Konformerengenerators (in diesem Falle OMEGA¹⁹⁹) spricht.

Bei Pharmakophor-Modellen, die nur nach energetischen Kriterien auf Basis von Ligand-Rezeptor-Komplexen erstellt werden (E-Pharmakophore, bzw. MIF basierte E-Pharmakophore), zeigte sich ein schlechteres Anreicherungsverhalten aktiver Moleküle als bei den mit CavKA oder Ligandscout erstellten Modellen. Dies ist vermutlich auf die *falsch-positive* Information der MIFs oder der Energie-Gitter der Docking Scoring-Funktion zurückzuführen (vgl. Kap. 19).

Protein-Ligand-Interaktions-Fingerprints (PLIFs) konnten erfolgreich dazu genutzt werden, um unterschiedliche Bindungsmodi von co-kristallisierten Liganden in unterschiedlichen PDE5-Ligand-Rezeptor-Komplexen zu identifizieren, die anschließend für die Erstellung von Pharmakophor-Modellen verwendet werden konnten. In einem *Parallel Screening* dieser Modelle und einer sich anschließenden Hitlisten-Vereinigung wurden erheblich bessere Ergebnisse als mit einem einzelnen Modell erzielt (s. Kap. 20).

Der Versuch, auf Basis von Ligand-Rezeptor-Komplexen strukturbasierte Gauß-Pharmakophore zu erstellen und so die Vorteile ligandzentrischer und strukturbasierter Ansätze zu kombinieren, zeigte, dass jene Hybrid-Modelle meist ein schlechteres Anreicherungsvermögen aufweisen als der rein ligandzentrische Ansatz.

23.2 Zusammenfassung ligandzentrische Methoden

Die in dieser Arbeit verglichenen ligandzentrischen Ansätze zeigten in vielen Fällen eine gute Leistungsfähigkeit, aktive Moleküle im Rahmen eines Virtuellen Screenings anzureichern. Ein großer Vorteil dieser Methoden ist, dass eine aufwändige Protein-Vorbereitung entfällt, zudem keine weitere Definition von Parametern nötig ist und somit der Einfluss des Anwenders auf die Screening-Ergebnisse (engl. *user bias*) extrem gering ist (s. Kap. 7).

Die besten Ergebnisse wurden mit Methoden erzielt, die nach wissensbasierten Ansätzen pharmakophore Eigenschaften definieren und diese zusätzlich mit der Information des Ligand-Volumens (engl. *shape*) in ihrer Bewertungsfunktion kombinieren. *Pharao* war das beste frei verfügbare Programm und zeigte verglichen mit *ROCS_{combo}* eine gute und vergleichbare Leistungsfähigkeit.

Methoden, die *shape* ($ROCS_{shape}$) und pharmakophore Eigenschaften regelbasiert definieren ($ROCS_{combo}$, *Pharao*), erwiesen sich als sehr robust gegenüber geringen Konformationsänderungen, sodass auch bei niedrig aufgelösten Konformer-Ensembles sehr gute Ergebnisse erzielt werden konnten. Bei Methoden, die ein elektrostatisches Potential zur Definition pharmakophorer Eigenschaften verwenden (EON_{combo} und $ShaEP_{combo}$), zeigte sich bei gleicher Konformer-Auflösung ein erheblich schlechteres Anreicherungsverhalten aktiver Moleküle.

Methoden, die Moleküle allein aufgrund ihres Volumens abgleichen ($ROCS_{shape}$, *Piramid*, $ShaEP_{shape}$), sind nicht in dem Maße in der Lage aktive Moleküle anzureichern, wie Methoden, die regelbasiert pharmakophore Eigenschaften in der Datenbanksuche berücksichtigen. *Shape*-basierte Ansätze erweisen sich gegenüber jenen, die elektrostatische Potentiale zur Definition pharmakophorer Eigenschaften verwenden, jedoch als überlegen.

Kapitel 24

Ausblick

Obwohl CavKA bereits in vielfacher Hinsicht optimiert wurde und sein Nutzen zur Erstellung strukturbasierter Pharmakophor-Modelle gezeigt werden konnte, ergeben sich dennoch einige Aspekte und Fragestellungen für zukünftige Arbeiten. Die Validierung der Ligand-Rezeptor-Interaktionserkennung in CavKA erfolgte auf einer relativ begrenzten Anzahl an Ligand-Rezeptor-Komplexen. Die Distanz- und Winkeltoleranzen sind deshalb bewusst großzügig gewählt worden. Eine Validierung der bestehenden Parameter, bzw. eine Re-Parametrisierung auf einem größeren Datensatz wäre empfehlenswert. Hierfür könnte der im November 2011 publizierte SERAPhiC-Datensatz verwendet werden, bei dem Ligand-Rezeptor-Interaktionen genau beschrieben sind³⁰⁹.

Die beste Möglichkeit, ein Programm ausgiebig zu validieren, ist ein ausgiebiger Beta-Test: Geplant ist, CavKA kostenlos und frei verfügbar der Allgemeinheit zur Verfügung zu stellen, um mit externen Testern das Programm stetig zu verbessern. Ferner bieten sich die Unterstützung weiterer Pharmakophor-Formate oder die Implementierung weiterer Pharmakophor-Definitionen an (wie z.B. projizierte Punkte oder alternative Definitionen von Hydrophobizität).

In seiner aktuellen Version ist CavKA ein in Python geschriebenes Kommandozeilen-Programm. Um die Bedienung noch weiter zu vereinfachen, bzw. um Ligand-Rezeptor-Interaktionen noch intuitiver darstellen zu können, wäre eine Einbindung in ein Programm zur dreidimensionalen Moleküldarstellung, das über Python als Programmiersprache angesteuert werden kann (z.B. PyMOL³¹⁰), eine sinnvolle Erweiterung.

Ein weiterer Aspekt ist der für die Validierung verwendete Fieldscreen-Datensatz, der massive Fehler und Schwächen aufweist. So sollte ein weiteres Augenmerk dem Design und der Entwicklung neuer Validierdatensätze gelten.

Teil IV

Anhang

Kapitel 25

Module und Programme

Nachfolgend wird die Integration von frei verfügbaren (engl. *open source*) Software-Modulen, bzw. deren jeweilige Verwendung in der CavKA-Methode kurz erläutert (vgl. Abschn. 25.1). Die Funktionen eigens implementierter Module wird in Abschnitt 25.3 umrissen. CavKA sowie die eigens implementierten Module sind in der frei verfügbaren Python-Programmiersprache geschrieben³¹¹.

Abbildung 25.1 zeigt die Verzahnung der besprochenen Module in der CavKA-Methode untereinander, bzw. mit den in Abschnitt 25.2 erläuterten Programmen. Ferner wird die Abhängigkeit von verwendeter frei verfügbarer Software gezeigt. Weitere in dieser Arbeit verwendete Programme sind in Abschnitt 25.4 kurz umrissen.

25.1 Open-Source-Module

25.1.1 Numpy

Numpy ist ein Python-Modul zur Prozessierung großer, mehrdimensionaler Vektoren³¹². Aufgrund der zugrundeliegenden C-Routinen sind numerische Operationen mit Numpy äußerst recheneffizient und performant.

25.1.2 hCluster

hCluster besitzt eine sehr leistungsfähige Funktion zur Berechnung von Distanzmatrizen³¹³. Das Modul wurde u.a. verwendet, um den Liganden umgebende MIF-Punkte zu detektieren und um die GRID-Felder auf das Bindetaschenareal zu begrenzen.

25.1.3 Biopython

Aus Biopython²²⁴ wurden überwiegend die Routinen Bio.PDB²²⁶ (für das Einlesen dreidimensionaler PDB-Röntgenkristallstrukturen) und Bio.KDtree (für Nachbarschaftssuchen von Ligandeigenschaften zu kompatiblen Aminosäuren der Binde-tasche) genutzt. Darüber hinaus bietet Biopython umfangreiche Programme zur Bearbeitung molekularbiologischer Fragestellungen²²⁵.

25.1.4 Openbabel und Pybel

Openbabel²¹⁷ ist ein Programm, das das Einlesen und Schreiben von unterschiedlichen Moleküldateiformaten ermöglicht sowie zur Dateikonvertierung oder zur Analyse von Molekülinformation (beispielsweise über bereits implementierte Fingerprints) genutzt werden kann. Pybel²²⁸ ist eine Schnittstelle, die die Verwendung von Openbabel-Routinen über Python ermöglicht.

25.1.5 Pharao

Pharao^{187,203} ist ein Open-Source-Projekt der belgischen Software Firma SILICOS NV. Der ligandenzentrische Ansatz annotiert pharmakophore Ligandeigenschaften über Openbabel-Routinen und bietet die Möglichkeit der Datenbanksuche.

25.1.6 Matplotlib und Pylab

Matplotlib ermöglicht das Erstellen sowie den Export hochwertiger 2D-Grafiken³¹⁴. Das Programm imitiert hierbei die Funktionalitäten der kommerziell erhältlichen Matlab®-Software in Bezug auf graphische Darstellungen und wurde zum Erstellen von ROC-Kurven und Heat-Maps genutzt. Pylab ist eine Schnittstelle, die die Verwendung von Matplotlib-Routinen direkt über Python ermöglicht.

25.2 Eigene CavKA-Programme

Nachfolgend werden die Aufgaben eigens implementierter Python-Programme sowie deren Integration in der CavKA-Methode näher erläutert. Jedes der genannten Programme importiert mindestens eines der unter Abschnitt 25.1, bzw. 25.3 erläuterten Module.

25.2.1 cavka.py

Dieses Programm nimmt den zentralen Punkt in der CavKA-Methode ein und übernimmt folgende Aufgaben:

- Definition pharmakophorer Ligandeigenschaften (vgl. Abschn. 10.1),
- Übergabe der Ligandeigenschaften an das Modul 'sift_cavka_pharao.py', wo die Ligandeigenschaften auf Interaktionen mit der Bindetasche analysiert werden (vgl. Abschn. 11),
- Übergabe der konturierten MIFs an das Modul cavka_E.py zur Ableitung energiebasierter Pharmakophore (vgl. Abschn. 13.2),
- Export der Pharmakophor-Modelle in die Formate .phar und .ph4 (vgl. Abschn. 10.3).

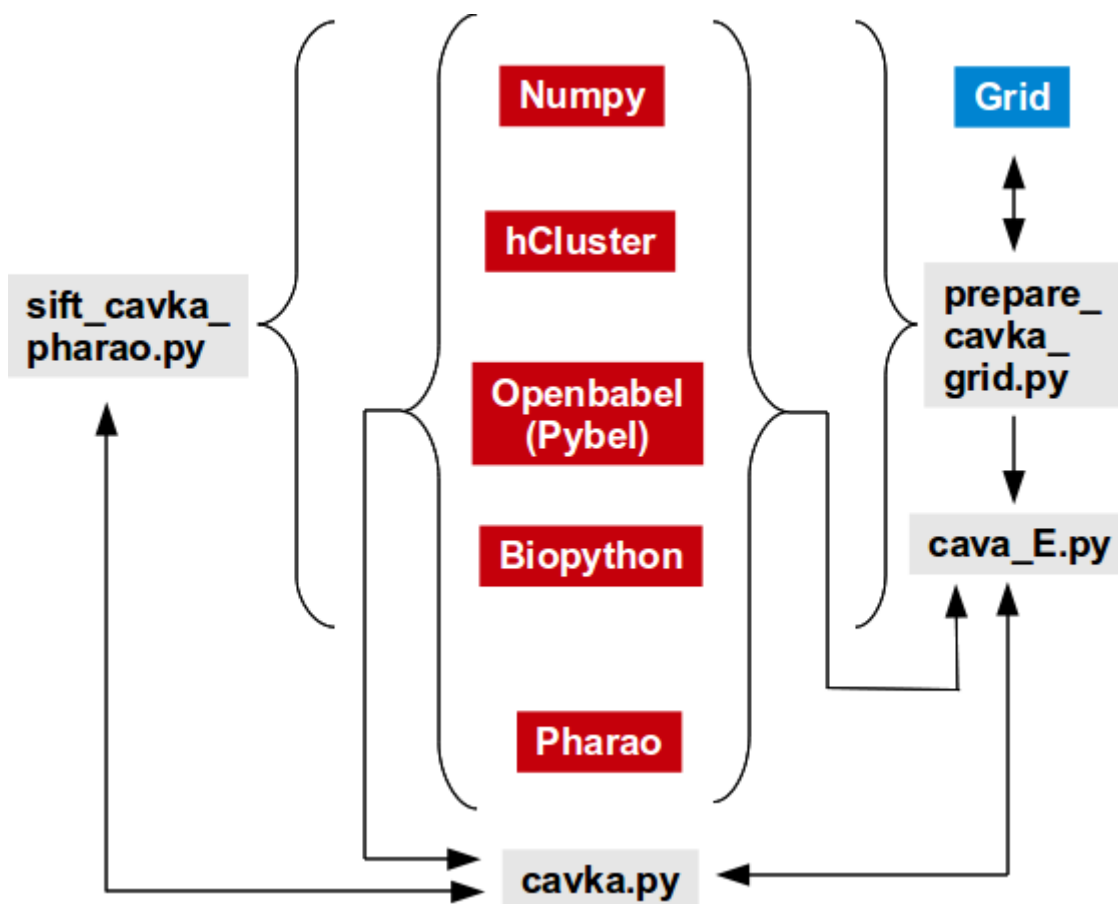


Abbildung 25.1: Übersicht über die Verknüpfung (Pfeile) von eigens implementierten CavKA-Modulen und-Programmen (grau) mit *Open-Source*-Modulen (rot), bzw. kommerziellen Programmen (blau). Weitere Erläuterungen s. Text.

25.2.2 `prepare_cavka_grid.py`

Mit dem Programm werden die Vorbereitungen für die Berechnungen der GRID-MIFs, bzw. die sich anschließenden MIF-Prozessierungen realisiert (vgl. Kap. 12). Es umfasst:

- Vorbereitung von PDB-Strukturen (konsistente Nummerierung von Aminosäure-Resten; Entfernung von Nicht-Standard-Aminosäuren und Wassermolekülen),
- Einlesen des Liganden und Dimensionierung der GRID-Gitterbox für die MIF-Berechnung,
- Ausführen der GRID-Berechnung,
- Reduktion des MIFs auf das Ligand-Areal,
- Normierung und Konturierung der MIFs,
- Ausgabe der erhaltenen *Hotspots* als .txt Dateien (für weitere Berechnungen) sowie als mol.2 Dateien (zur Visualisierung).

25.3 Eigene CavKA-Module

25.3.1 `sift_cavka_pharao.py`

In diesem Modul sind die geometrischen Regeln für zustande kommende Ligand-Rezeptor-Interaktionen (vgl. Kap. 11) sowie die pharmakophoren Eigenschaften von Aminosäuren der Bindetasche definiert (vgl. Abschn. 10.2). Die zuvor über das Programm 'cavka.py' (vgl. Abschn. 25.2.1) ermittelten pharmakophoren Ligandeigenschaften (vgl. Abschn. 25.2.1) werden an das Modul übergeben und auf kompatible Ligand-Rezeptor-Interaktionen hin analysiert.

25.3.2 `cavka_E.py`

An dieses Modul werden die skalierten und konturierten MIFs übergeben, die mit den zuvor über das Programm 'cavka.py' (vgl. Abschn. 25.2.1) annotierten Ligandeigenschaften abgeglichen werden. Nach den in Abschnitt 13.2 erläuterten Regeln werden Ligandeigenschaften über die durchschnittlichen MIF-Interaktionswerte priorisiert, an das 'cavka.py' Programm zurückgegeben und schließlich in ein MIF-basiertes E-Pharmakophor exportiert (vgl. auch Abb. 25.1).

25.4 Weitere kommerzielle, bzw. frei verfügbare Programme

25.4.1 PyMOL

Pymol ist ein Programm zur dreidimensionalen Visualisierung von Proteinen und Molekülen³¹⁰. Neben zahlreichen Datei-Formaten unterstützt PyMOL *Ray-Tracing*, was den Export hochqualitativer Bilder ermöglicht. Sofern nicht anders angemerkt, sind alle dreidimensionalen Darstellungen von Molekülen oder Proteinen in der vorliegenden Arbeit mit Pymol erstellt worden.

25.4.2 MOE

Das kommerziell von CCG vertriebene Programm MOE bietet Anwendungen für nahezu jede erdenkliche Aufgabenstellung im *Molecular-Modelling*⁶². In der vorliegenden Arbeit wurde MOE zur Editierung von CavKA-Pharmakophor-Modellen im .ph4 Format, zur Datenbanksuche, für die Superpositionierung von Ligand-Rezeptor-Komplexen, für PLIF-Berechnungen, für *Multi Fragment Search* (MFSS) und für die Optimierung von Ligand-Rezeptor-Komplexen über *Protonate-3D*²²² verwendet.

25.4.3 GRID

Das kommerziell von Molecular Discovery vertriebene Programm GRID wurde zur Berechnung der Molekularen Interaktions-Felder (Abk. MIFs) verwendet²³².

Kapitel 26

Quellcode

26.1 Modul sift_cavka.py

```
1  ### sift_cavka.py
2  ### 09-may-2011
3  ### --> changed !!
4  ### using pharao core!
5  ### DEBUGGING history
6  ### 15-mai-2011(fk)
7  ### DEBUGGING OH Donors; avoid using normals (adding noise)
8  ### 16-may-2011(fk)
9  ### DEBUGGING 0 Acceptor normals; 1A too far away from heavy atom;
10 ### --> resulting in wrong angles
11 ### 20-may-2011(fk)
12 ### implenting Energetic Pharmacophores:
13 ### - check grid potential at feature
14 ### - norm energy !
15 ### - keep the best energy Pharmacophores and use them for final Hypothesis
16 ### 25-may-2011(fk)
17 ### reduce alpha for EXCL! 1.7 A is too big!
18
19 from time import*
20 import sys,os,time
21 import pybel
22 import openbabel
23 import math
24 import numpy
25 from numpy import *
26
27 ### using BIO.PDB 1.51 -- uses numpy!!; since 21-aug-09; important for the use of calc_dihedral
28 ### used MODULE : Bio.PDB.Vector
29 from Bio.PDB import *
30
31 #####
32 ### Rules 4 SIFT ###
33 #####
34
35 ### ATTENTION TAKE CARE 4 COMPLEMENTARITY!!!
36 ### Ionical definitions for further implementations
37 ### 1st part: all donors from sidechain
38 ### GLU_OE 2 donor
39 dict_donor={'ARG_NE': ['HE'], 'ARG_NH1': ['1HH1', '2HH1'], 'ARG_NH2': ['1HH2', '2HH2'], 'ASN_ND2': ['1HD2',
40 '2HD2'], 'GLN_NE2': ['1HE2', '2HE2'], 'HIS_NE2': ['2HE'], 'LYS_NZ': ['1HZ', '2HZ', '3HZ'], 'SER_OG': ['
41 'HG'], 'THR_OG1': ['1HG'], 'TRP_NE1': ['1HE'], 'TYR_OH': ['HH'],
42
43 ### 2nd part: all entries from Main chain
44 'ALA_N': ['H'], 'ARG_N': ['H'], 'ASN_N': ['H'], 'ASP_N': ['H'], 'CYS_N': ['H'], 'GLN_N': ['H'], 'GLU_N': ['H',
45 'GLY_N': ['H'], 'HIS_N': ['H'], 'ILE_N': ['H'], 'LEU_N': ['H'], 'LYS_N': ['H'], 'MET_N': ['H'], '
46 PHE_N': ['H'], 'PRO_N': ['H'], 'SER_N': ['H'], 'THR_N': ['H'], 'TRP_N': ['H'], 'TYR_N': ['H'], 'VAL_N'
47 : ['H']}
48
49 ### 09-may-2011 (fk)
```

```

43 ### POSITIVE CHARGED CENTER
44 ### 'LYS_NZ', 'ARG_NE', 'HIS_NE2' --> cationic
45 ### consider all atoms as (possibly) charged
46 ### 18-may-2011 (fk) metal interactions;
47 POSC_rules = [ 'LYS_NZ', 'ARG_NE', 'HIS_NE2' ]
48 ### 25-aug-09 (fk)
49 ### what about AS NHX Sidechains (acting as acceptors???)
50 ### new acceptor dict for acceptor atoms and neighbor atoms --> advanced angle checking
51 dict_acceptor={ 'TYR_OH': [ 'CZ' ], 'SER_OG': [ 'CB' ], 'THR_OG1': [ 'CB' ], 'ASP_OD1': [ 'CG' ], 'GLU_OE1': [ 'CD' ],
52               GLN_OE1': [ 'CD' ], 'ASN_OD1': [ 'CG' ], 'CYS_SG': [ 'CB' ],
53 ## 2nd part: main chain heavy atoms
54 'ALA_O': [ 'C' ], 'ARG_O': [ 'C' ], 'ASN_O': [ 'C' ], 'ASP_O': [ 'C' ], 'CYS_O': [ 'C' ], 'GLN_O': [ 'C' ], 'GLU_O': [ 'C' ],
55 'GLY_O': [ 'C' ], 'HIS_O': [ 'C' ], 'ILE_O': [ 'C' ], 'LEU_O': [ 'C' ], 'LYS_O': [ 'C' ], 'MET_O': [ 'C' ], 'PHE_O': [ 'C' ],
56 'PRO_O': [ 'C' ], 'SER_O': [ 'C' ], 'THR_O': [ 'C' ], 'TRP_O': [ 'C' ], 'TYR_O': [ 'C' ], 'VAL_O': [ 'C' ], 'ASP_OD2': [ 'CG' ], 'GLU_OE2': [ 'CD' ] }
57 ### 10-may-2011 (fk)
58 ### ASP / GLU --> NEGATIVE
59 NEGC_rules= [ 'GLU_OE1', 'GLU_OE2', 'GLU_CD', 'ASP_CG', 'ASP_OD1', 'ASP_OD2' ]
60 ###
61 ### Hydrophobic definitions
62 SDRY= [ 'LEU_CB', 'LEU_CG', 'LEU_CD1', 'LEU_CD2', 'PRO_CB', 'PRO_CG', 'PRO_CD', 'TYR_CB', 'TYR_CG', 'TYR_CD1', 'TYR_CD2', 'TYR_CE1', 'TYR_CE2', 'TYR_CZ', 'LYS_CB', 'LYS_CG', 'LYS_CD', 'LYS_CE', 'TRP_CB', 'TRP_CG', 'TRP_CD1', 'TRP_CD2', 'TRP_CE2', 'TRP_CE3', 'TRP_CZ2', 'TRP_CZ3', 'TRP_CH2', 'PHE_CB', 'PHE_CG', 'PHE_CD1', 'PHE_CD2', 'PHE_CE1', 'PHE_CE2', 'PHE_CZ', 'ARG_CB', 'ARG_CG', 'ARG_CD', 'VAL_CB', 'VAL_CG1', 'VAL_CG2', 'ILE_CB', 'ILE_CG1', 'ILE_CG2', 'ILE_CD1', 'MET_CB', 'MET_CG', 'MET_SD', 'MET_CE' ]
63 ### 13-sep-2010
64 ### Aromatic definitions for pi interactions
65 AROMATIC= [ 'PHE', 'TRP', 'HIS', 'TYR' ]
66
67 def processing_protein(protein):
68     ### Dict -> keys: res_id / val: res_name
69     resdict={}
70     ### needs 2 read in complete file!
71     try:
72         structure = PDBParser().get_structure('X', protein)
73         structure = structure[0]
74         for chain in structure:
75             for residue in chain:
76                 res_id = residue.get_id()
77                 res_nr = res_id[1] ### Number of Res ID
78                 res_name = residue.get_resname()
79                 resdict[res_nr] = res_name
80     except Exception, e:
81         error = [None, None, "structure import error: %s" % e, protein]
82         print error
83         raise e
84 ### very simple function for reading out atom coords of a mol file
85 def get_lig_coords_mol(ligand_mol):
86     print "\n"
87     print "Started reading out ligand mol_coords"
88     print "\n"
89     mol = pybel.readfile("mol", ligand_mol)
90     pre_lig_coord_array = []
91     for element in mol:
92         for atom in element:
93             atom_coords = atom.coords
94             pre_lig_coord_array.append(atom_coords)
95     lig_coord_array = asarray(pre_lig_coord_array)
96     ### return array; iteration over the objects still possible !!!
97     return lig_coord_array
98
99 def dict_the_lig(ref_ligand):
100     ### 2dicts --> 1) coord: idx; 2) idx : coord
101     coord_2_idx = {}
102     idx_2_coord = {}
103     ###13-mrz-09 read in as mol
104     ### ALL LIGAND COORDS WITH INDEX --> IMPORTANT WHEN REINDEXING
105     ref_lig_mol = pybel.readfile("mol", ref_ligand)
106     for element in ref_lig_mol:
107         for atom in element:
108             atom_coords = atom.coords
109             atom_idx = atom.idx
110             entry = {atom_coords:atom_idx} ##1
111             coord_2_idx.update(entry)
112             entry_2 = {atom_idx: atom_coords} ##2

```

```

111         idx_2_coord.update(entry_2)
112     print "created lig - dict !"
113     print "\n"
114     return coord_2_idx, idx_2_coord
115
116     ### 09-may-2011 (fk)
117 def create_EXCL(structure ,lig_coords, alpha):
118     print "creating excluded volume ..."
119     print "\n"
120     EXCL = []
121     for center in lig_coords:
122         ### Neighbour search!
123         query_atoms = Selection.unfold_entities(structure,'A') ### Seraching complete Structure!!
124         ns = NeighborSearch(query_atoms)
125         ### CENTER --> UPPER DIST RANGE
126         neighbors = ns.search(center,6.0)
127         #print neighbors
128         try:
129             atom_list = Selection.unfold_entities(neighbors,'A')###Atom List containing ALL neighbours!
130             for atom in atom_list:
131                 #####
132                 ###--> DISCLAIM H-ATOMS !!!! 01sept2011 ###
133                 #####
134                 ###getting the Parent AS from ALL atoms within radius
135                 coords = atom.get_coord()
136                 x = coords[0]
137                 y = coords[1]
138                 z = coords[2]
139                 entry = [x,y,z,alpha,0,0,0,0]
140
141                 if entry not in EXCL:
142                     EXCL.append(entry)
143                 else:
144                     print "duplicate"
145                     print "\n"
146             except:
147                 print "No neighbors found within range !"
148                 print "\n"
149     EXCL.sort()
150     return EXCL
151
152     ### separate function 4 Dry
153     ### generic hydrophobic interactions LIPO / HYBL Lipophilic region
154 def sift_HYD(HYD, structure, rules, RANGE, Min_contact_res):
155     print "start checking lipophilic interactions ..."
156     print "\n"
157     HYD_interact = []
158     for element in HYD:
159         center = asarray([element[0], element[1], element[2]])
160         ### Neighbour search!
161         query_atoms = Selection.unfold_entities(structure,'A')
162         ### Seraching complete Structure!!
163         ns=NeighborSearch(query_atoms)
164         ### CENTER --> UPPER DIST RANGE
165         neighbors = ns.search(center,RANGE[-1])
166         try:
167             atom_list=Selection.unfold_entities(neighbors,'A') ### Atom List containing ALL neighbours
168             !
169             ### Atoms which make contact --> + getting Parents // entire residue
170             ### List with all RESIDUES that passed rules
171             reg_list = [] ### one Reg List 4 every feature!
172             for atom in atom_list:
173                 ###getting the Parent AS from ALL atoms within radius
174                 res = atom.get_parent()
175                 res_name = res.get_resname()
176                 atom_id = atom.get_id()
177                 ### Matching the right AS atom
178                 match_str = str(res_name)+"_"+(atom_id)
179                 #####
180                 ### DRY Contacts --> only Sidechain ###
181                 #####
182                 ### Exclusion 1: right AS Atom
183                 ### new list 4 all who passed reg Expression
184                 ###--> Hydrophobic Atoms
185                 if match_str in SDRY:
186                     reg_list.append(atom)
187             ### ensure having at least 3 hydrophobic contacts

```

```

187     if len(reg_list) >= Min_contact_res:
188         for reg_atom in reg_list:
189             reg_atom_id = reg_atom.get_id()
190             ### getting the Parent AS from ALL atoms within radius
191             reg_res = reg_atom.get_parent()
192             reg_res_name = reg_res.get_resname()
193             reg_res_id = reg_res.get_id()[1]
194             ### atom coords as vector
195             reg_atom_coords = numpy.asarray(reg_atom.get_coord())
196             ### Calc Dist - using numpy arrays for both coords!
197             dist = math.sqrt(sum((center-reg_atom_coords)**2))
198             ### Exclusion 2: above clash distance
199             ### LOWER DIST RANGE
200             if dist >= RANGE[0]:
201                 ### appending 2 filtered SIFT list + removing Doubles
202                 if element not in HYD_interact:
203                     HYD_interact.append(element)
204                 else:
205                     print "Double Entry -- not appending to list"
206                     AS_str = str(reg_res_name)+" "+str(reg_res_id)+" "+str(reg_atom_id)+" "+str(
207                         reg_atom_coords)+" "+str(dist)+" Dist " +" MATCH SIDECHAIN"
208                     print AS_str
209                     AS_str = match_str+" "+str(reg_res_id)+" "+str(reg_atom_coords)+" "+str(dist)+"
210                         Dist "+" FUNCTIONAL GROUP CLASH! SIDECHAIN"
211                     print AS_str
212             ### score 2 low / not enough contacts
213             else:
214                 print "Not enough hydrophobic contacts"
215                 print "\n"
216             ##### no neighbours
217             except:
218                 print "No neighbours 4 DRY Sphere found within tolerance sphere!"
219                 no_neighbours_str = " NO NEIGHBOURS 4 DRY WITHIN SEARCH RADIUS "
220             return HYD_interact
221
222     ### updated SIFT; using DON-H-ACC and H-ACC-X angle criterions; 07-sept-09 (fk)
223     ### HDON and HYBH !
224     ### 17-may-2011 (fk)
225     ### ignoring pharao normals (focussing on H-DON)
226     ### Fuzzy angles for NEGC - H interactions (mod xC-ACC-H angle)
227     def sift_HDON(HDON, structure, coord_2_idx, idx_2_coord, ref_ligand, rules, rules_II, RANGE):
228         print "\n"
229         print "start checking hdonor interactions "
230         print "\n"
231         ### 17-may-2011 (fk)
232         ### DON- H only !
233         ### filling up heavy DON idx
234         heavy_donor_idx = []
235         for element in HDON:
236             heavy_donor_coords = (element[0], element[1], element[2])
237             heavy_d_idx = coord_2_idx[heavy_donor_coords]
238             heavy_donor_idx.append(heavy_d_idx)
239         ### final dict
240         heavy_don_and_H = {}
241         ### reading in (again) ligand.mol via Openbabel --> for bond DETECTION!!!
242         obConversion = openbabel.OBConversion()
243         obConversion.SetInAndOutFormats("mol", "mol")
244         mol = openbabel.OBMol()
245         obConversion.ReadFile(mol, ref_ligand)
246         ### DEBUGGING DONORS
247         molecule = openbabel.OBMol()
248         for element in heavy_donor_idx:
249             ### via index --> relevant coords on ligand
250             nu_donor = mol.GetAtom(element)
251             ### from idx back 2 coords
252             heavy_don_coords = idx_2_coord[element]
253             ### summarized list; all H (potential) H atoms next 2 the heavy donor
254             pot_H_coords = []
255             ### Openbabel: Iter Object has to be generated!
256             for atom in openbabel.OBMolAtomIter(mol):
257                 ### 17-mrz-09: parsing mol reveals no direct elements -> use atomic numbers
258                 ### check for all atoms connected 2 heavy_atom
259                 check = nu_donor.IsConnected(atom)
260                 if check == True:
261                     #####if H ia connected...
262                     if atom.GetAtomicNum() == 1:

```

```

262         coords = [atom.GetX(), atom.GetY(), atom.GetZ()]
263         H_coord_list = coords
264         pot_H_coords.append(H_coord_list)
265         if len(pot_H_coords) > 0:
266             don_entry = {heavy_don_coords: pot_H_coords}
267             heavy_don_and_H.update(don_entry)
268     HDON_interact = []
269     for element in HDON:
270         query = (element[0], element[1], element[2])
271         heavy_don_H = (heavy_don_and_H[query])
272         print "\n"
273         print heavy_don_H
274     ### conventional interaction unit !!!
275     ### vis center
276     don_center = asarray([element[0], element[1], element[2]])
277     x = molecule.NewAtom()
278     x.SetAtomicNum(7)
279     x.SetVector(float(element[0]), float(element[1]), float(element[2]))
280     ### vis norm
281     don_norm = asarray([element[5], element[6], element[7]])
282     x = molecule.NewAtom()
283     x.SetAtomicNum(12)
284     x.SetVector(float(don_norm[0]), float(don_norm[1]), float(don_norm[2]))
285     ### vis attached heavy don H
286     for i in heavy_don_H:
287         print i
288         x = molecule.NewAtom()
289         x.SetAtomicNum(1)
290         x.SetVector(float(i[0]), float(i[1]), float(i[2]))
291     ### Neighbour search!
292     query_atoms = Selection.unfold_entities(structure, 'A') ### Seraching complete Structure!!
293     ns = NeighborSearch(query_atoms)
294     ### Searching to upper Limit! --> ACC on porotein Site
295     neighbors = ns.search(don_center, RANGE[-1]) ### upper dist range!
296     print neighbors
297     try:
298         atom_list = Selection.unfold_entities(neighbors, 'A') ### Atom List containing ALL neighbours
299         !
300         ### Atoms which make contact --> + getting Parents // entire residue
301         for atom in atom_list:
302             atom_id = atom.get_id()
303             ### getting the Parent AS from atom within radius
304             res = atom.get_parent()
305             res_name = res.get_resname()
306             res_id = res.get_id()[1]
307             ### atom coords as array
308             atom_coords = numpy.asarray(atom.get_coord())
309             ### Matching the right AS atom
310             match_str = str(res_name) + "-" + str(atom_id)
311             dist = math.sqrt(sum((don_center - atom_coords)**2))
312             ### rules is now a dict!!! ; 07-sep-09 (fk)
313             if match_str in rules:
314                 ### Exclusion 2: above clash distance
315                 if dist >= RANGE[0]:
316                     ### collecting all coords 4 angle calc ###
317                     ### protein acceptor
318                     heavy_site_acceptor = numpy.array(atom_coords)
319                     x = molecule.NewAtom()
320                     x.SetAtomicNum(8)
321                     x.SetVector(float(heavy_site_acceptor[0]), float(heavy_site_acceptor[1]), float(
322                         heavy_site_acceptor[2]))
323                     ### angle calculation I (ideally 180 degrees)
324                     ### DON - H - Acc
325                     print heavy_don_H
326                     for H in heavy_don_H:
327                         vec_1 = heavy_site_acceptor - asarray(H)
328                         vec_2 = don_center - asarray(H)
329                         skalar = numpy.inner(vec_1, vec_2)
330                         norm_1 = math.sqrt(sum((vec_1)**2))
331                         norm_2 = math.sqrt(sum((vec_2)**2))
332                         angle = math.degrees(math.acos((skalar)/(norm_1*norm_2)))
333                         ### correct angles
334                         if 110.0 < angle < 180.0:
335                             print "\n"
336                             print angle, match_str, res_id, " ACC-H-Don PASSED ANGLE SIFT DONOR"
337                             print "\n"
338                             ### 07-sep-09 (fk)

```

```

337     ### checking C-O-H Angle
338     acc_C_id = (rules[match_str])[0]
339     ### get C_Atom Coords
340     C_acceptor = res[acc_C_id]
341     C_acceptor_coords = numpy.asarray(C_acceptor.get_coord())
342     x = molecule.NewAtom()
343     x.SetAtomicNum(8)
344     x.SetVector(float(C_acceptor_coords[0]), float(C_acceptor_coords[1]), float(
        C_acceptor_coords[2]))
345     ### Cx-Acc-H (ideally 120 degrees)
346     vec_3 = C_acceptor_coords-heavy_site_acceptor
347     vec_4 = asarray(H)-heavy_site_acceptor
348     skalar_II = numpy.inner(vec_3, vec_4)
349     norm_3 = math.sqrt(sum((vec_3)**2))
350     norm_4 = math.sqrt(sum((vec_4)**2))
351     angle_II = math.degrees(math.acos((skalar_II)/(norm_3*norm_4)))
352     ### changed for serine proteases; 25-jan-10
353     if 90.0 < angle_II < 170.0:
354         ### final appending
355         if element not in HDON_interact:
356             HDON_interact.append(element)
357             AS_str = match_str+" "+str(res_id)+" "+str(atom_coords)+" "+str( "%.3f" %
                dist) +" DIST "+str("%.3f" % angle)+ " ANGLE " + " MATCHING H ATOM "
358     ### wrong angles
359     else:
360         print "WRONG ANGLES"
361         ### except NEGC - DONOR interactions
362         if match_str in rules_II:
363             #print "MATCH STR ii", match_str
364             #print "\n"
365             if 55.0 < angle_II < 100.0:
366                 print angle_II, 'passed (NEGC EXCEPTION)'
367                 print "\n"
368                 if element not in HDON_interact:
369                     HDON_interact.append(element)
370                     print "excepted NEGC -- Donor Interaction"
371     ### clash distance
372     else:
373         AS_str = match_str+" "+str(res_id)+" "+str(atom_coords)+" "+str(dist)+" Dist "+"
            FUNCTIONAL GROUP CLASH! MAIN CHAIN "
374         print AS_str
375     ### wrong matching str
376     else:
377         print "match_string not in rules"
378     else:
379         print "\n"
380         print "Checking HDON_HAcc Interactions - Contacts !! "
381         print "\n"
382     except:
383         print "No neighbours found within tolerance sphere!"
384         no_neighbours_str = " NO NEIGHBOURS WITHIN SEARCH RADIUS "
385     pybelmol = pybel.Molecule(molecule)
386     pybelmol.write("sdf", 'donor_DEBUG.sdf', overwrite = True)
387     print 'interactions successfully written to sdf!'
388     print "\n"
389     ### display interacting donors !
390     interact_mol= openbabel.OBMol()
391     for i in HDON_interact:
392         print i, 'PASSED DONOR INTERACTION'
393         print i[0], i[1], i[2]
394         x = interact_mol.NewAtom()
395         x.SetAtomicNum(7)
396         x.SetVector(float(i[0]), float(i[1]), float(i[2]))
397     pybelmol_II = pybel.Molecule(interact_mol)
398     pybelmol_II.write("sdf", 'donor_INTERACT.sdf', overwrite = True)
399     return HDON_interact
400
401     ### 11-sept-2011 --> checking PI-PI (edge 2 face; face 2 face)
402     ### PDB 1lpz example (edge 2 face)
403     def sift_PI_PI(AROM, structure, rules, RANGE):
404         print "\n"
405         print 'started checking pi-pi interactions'
406         print "\n"
407         PI_PI_interact = []
408         ### write final interact mol
409         mol_interact = openbabel.OBMol()
410         ### PHE_local = ['CG', 'CZ', 'CD1', 'CD2', 'CE1', 'CE2'] ### PHE / TYR --> IDENTICAL

```

```

411 TRP_local_I = ['CE2','CD2','CE3','CZ3','CH2','CZ2']### TRP big ring
412 TRP_local_II = ['CE2','CD2','NE1','CD1','CG'] ### TRP small ring
413 ### HIS_local = ['CE1','CD2','CG','ND1','NE2']
414 AROMATIC_local = ['CG','CZ','CD1','CD2','CE1','CE2',
415                  'CE1','CD2','CG','ND1','NE2']
416 ### debug
417 mol = openbabel.OBMol()
418 for element in AROM:
419     center = asarray([element[0],element[1],element[2]])
420     ### Neighbour search!
421     query_atoms=Selection.unfold_entities(structure,'A') ### Seraching complete Structure!!
422     ns=NeighborSearch(query_atoms)
423     res_control= []
424     ### CENTER --> UPPER DIST RANGE
425     neighbors = ns.search(center,RANGE[-1])
426     try:
427         prot_planes = []
428         atom_list=Selection.unfold_entities(neighbors,'A') ### Atom List containing ALL neighbours
429         !
430         ### Atoms which make contact --> + getting Parents // entire residue
431         for atom in atom_list:
432             atom_id = atom.get_id()
433             ###getting the Parent AS from atom within radius; eventually several atoms
434             ###eventually several atoms belonging to the same residue!
435             res = atom.get_parent()
436             chain = res.get_parent()
437             res_name = res.get_resname()
438             res_id = res.get_id()[1]
439             check_str = res_name+'_'+str(res_id)
440
441             ### Exclusion 1: right AS Atom --> if one residue is matching; check whole res!
442             if res_name in rules and res_id not in res_control:
443                 res_control.append(res_id)
444
445             ### for all ring atoms to check!####
446             hit_list = []
447             hit_list_II = []
448             ### coords for AS protein planes #####
449             if res_name != 'TRP':
450                 hit_res = chain[res_id]
451                 print hit_res
452                 for hit_atom in hit_res:
453                     hit_id = hit_atom.get_id()
454                     if hit_id in AROMATIC_local:
455                         print hit_id
456                         hit_coords = hit_atom.get_coord()
457                         entry = [hit_coords[0],hit_coords[1], hit_coords[2]]
458                         hit_list.append(entry)
459                 print asarray(hit_list)
460                 print "\n"
461                 ### ring centers
462                 mean_point = (asarray(hit_list)).mean(axis=0)
463                 #print mean_point
464                 entry_mean = [mean_point[0],mean_point[1],mean_point[2]]
465                 a = hit_list[0]
466                 b = hit_list[1] ### HIS: only 5 members!
467
468                 ### normal vector ...
469                 ### 17-sep-2011
470                 vec_a = a - mean_point
471                 vec_b = b - mean_point
472                 to_proj_a = cross(vec_a, vec_b)
473                 to_proj_b = cross(vec_b, vec_a)
474                 norm_a = math.sqrt(sum((to_proj_a)**2))
475                 norm_b = math.sqrt(sum((to_proj_b)**2))
476                 ### Normalize projected points !!!
477                 ### 2.1A --> analogue to moe proj pts
478                 new_a = (to_proj_a/ norm_a)*2.1
479                 new_b = (to_proj_b/ norm_b)*2.1
480                 ### adding to ring center
481                 proj_pt_a = new_a + mean_point
482                 proj_pt_b = new_b + mean_point
483                 plane_entry = [entry_mean, proj_pt_a, proj_pt_b]
484                 prot_planes.append(plane_entry)
485             ### except TRP!
486             if res_name == 'TRP':
487                 hit_res = chain[res_id]

```

```

487     for hit_atom in hit_res:
488         hit_id = hit_atom.get_id()
489         ### TRP big ring
490         if hit_id in TRP_local_I:
491             hit_coords = hit_atom.get_coord()
492             entry_I = [hit_coords[0], hit_coords[1], hit_coords[2]]
493             hit_list.append(entry_I)
494         ### TRP small ring
495         if hit_id in TRP_local_II:
496             hit_coords = hit_atom.get_coord()
497             entry_II = [hit_coords[0], hit_coords[1], hit_coords[2]]
498             hit_list_II.append(entry_II)
499     ### Process big ring
500     mean_point_I = (asarray(hit_list)).mean(axis=0)
501     entry_mean_I = [mean_point_I[0], mean_point_I[1], mean_point_I[2]]
502     a = hit_list[0]
503     b = hit_list[1] ### HIS: only 5 members!
504     ### normal vector ...
505     ### 17-sep-2011
506     vec_a = a - mean_point_I
507     vec_b = b - mean_point_I
508     to_proj_a = cross(vec_a, vec_b)
509     to_proj_b = cross(vec_b, vec_a)
510     norm_a = math.sqrt(sum((to_proj_a)**2))
511     norm_b = math.sqrt(sum((to_proj_b)**2))
512
513     ### Normalize projected points !!!
514     ### 2.1A --> analogue to moe proj pts
515     new_a = (to_proj_a/ norm_a)*2.1
516     new_b = (to_proj_b/ norm_b)*2.1
517     ### adding to ring center
518     a_I = new_a + mean_point_I
519     b_I = new_b + mean_point_I
520     ### AS_plane
521     plane_entry_I = [entry_mean_I, a_I, b_I]
522     prot_planes.append(plane_entry_I)
523     ### Process small ring
524     mean_point_II = (asarray(hit_list_II)).mean(axis=0)
525     entry_mean_II = [mean_point_II[0], mean_point_II[1], mean_point_II[2]]
526     a_II = hit_list_II[0]
527     b_II = hit_list_II[1] ### HIS: only 5 members!
528     ### normal vector ...
529     ### 17-sep-2011
530     vec_a_II = a_II - mean_point_II
531     vec_b_II = b_II - mean_point_II
532     to_proj_a_II = cross(vec_a_II, vec_b_II)
533     to_proj_b_II = cross(vec_b_II, vec_a_II)
534     norm_a_II = math.sqrt(sum((to_proj_a_II)**2))
535     norm_b_II = math.sqrt(sum((to_proj_b_II)**2))
536     ### Normalize projected points !!!
537     ### 2.1A --> analogue to moe proj pts
538     new_a_II = (to_proj_a_II/ norm_a_II)*2.1
539     new_b_II = (to_proj_b_II/ norm_b_II)*2.1
540     ### adding to ring center
541     a_II = new_a_II + mean_point_II
542     b_II = new_b_II + mean_point_II
543     ### AS_plane
544     plane_entry_II = [entry_mean_II, a_II, b_II]
545     prot_planes.append(plane_entry_II)
546     ### LIGAND ...
547     p1 = asarray([element[5], element[6], element[7]]) - center
548     norm_p1 = math.sqrt(sum((p1)**2))
549     ### 2.1 A moe - alike
550     new_p1 = (p1/ norm_p1)*2.1
551     ### upper and lower ring plane
552     proj_p1 = center + new_p1
553     proj_p2 = center - new_p1
554     ### DEBUG LIGAND...
555     y = mol.NewAtom()
556     y.SetAtomicNum(9)
557     y.SetVector(float(element[0]), float(element[1]), float(element[2]))
558     p1 = mol.NewAtom()
559     p1.SetAtomicNum(10)
560     p1.SetVector(float(proj_p1[0]), float(proj_p1[1]), float(proj_p1[2]))
561
562     p2 = mol.NewAtom()
563     p2.SetAtomicNum(10)

```

```

564 p2.SetVector(float(proj_p2[0]),float(proj_p2[1]),float(proj_p2[2]))
565 ### check angles between prot_vec and lig_vec
566 ### prot_planes = prot_planes[2:3]
567 if len(prot_planes)> 0:
568     ### 1st center; 2nd/3rd proj. pts.
569     ### closer proj pt from protein
570     for i in prot_planes:
571         center_dist = math.sqrt(sum((center-i[0])**2))
572         if center_dist >= RANGE[0] and center_dist <= RANGE[-1]:
573             a = math.sqrt(sum((center-i[1])**2))
574             b = math.sqrt(sum((center-i[2])**2))
575             if a < b:
576                 proj_PROT = i[1]
577             else:
578                 proj_PROT = i[2]
579             ### closer proj pt from ligand
580             c = math.sqrt(sum((proj_p1-i[0])**2))
581             d = math.sqrt(sum((proj_p2-i[0])**2))
582             if c < d:
583                 proj_LIG = proj_p1
584             else:
585                 proj_LIG = proj_p2
586             ### EDGE 2 FACE
587             ### alpha: proj_Lig - center_Lig - center_Prot ### EDGE 2 FACE
588             vec_1 = proj_LIG - center
589             vec_2 = i[0] - center
590             skalar = numpy.inner(vec_1, vec_2)
591             norm_1 = math.sqrt(sum((vec_1)**2))
592             norm_2 = math.sqrt(sum((vec_2)**2))
593             ### (more restrictive)
594             alpha = math.degrees(math.acos((skalar)/(norm_1*norm_2)))
595             if alpha >180.0:
596                 alpha = alpha - 180.0
597             ### EDGE 2 FACE and FACE 2 FACE
598             ### beta : center_Prot - proj_Prot - center_lig
599             vec_3 = center - proj_PROT
600             vec_4 = i[0] - proj_PROT
601             skalar = numpy.inner(vec_3, vec_4)
602             norm_3 = math.sqrt(sum((vec_3)**2))
603             norm_4 = math.sqrt(sum((vec_4)**2))
604             beta = math.degrees(math.acos((skalar)/(norm_3*norm_4)))
605             if beta >180.0:
606                 beta = beta - 180.0
607             ### FACE 2 FACE
608             ### beta : center_LIG - proj_LIG - center_PROT
609             vec_5 = center - proj_LIG
610             vec_6 = i[0] - proj_LIG
611             skalar = numpy.inner(vec_5, vec_6)
612             norm_5 = math.sqrt(sum((vec_5)**2))
613             norm_6 = math.sqrt(sum((vec_6)**2))
614             gamma = math.degrees(math.acos((skalar)/(norm_5*norm_6)))
615             if gamma >180.0:
616                 gamma = gamma - 180.0
617             ###-----
618             p1 = mol.NewAtom()
619             p1.SetAtomicNum(9)
620             p1.SetVector(float((i[0])[0]),float((i[0])[1]),float((i[0])[2]))
621             p2 = mol.NewAtom()
622             p2.SetAtomicNum(10)
623             p2.SetVector(float(proj_LIG[0]),float(proj_LIG[1]),float(proj_LIG[2]))
624             p3 = mol.NewAtom()
625             p3.SetAtomicNum(9)
626             p3.SetVector(float(center[0]),float(center[1]),float(center[2]))
627             p4 = mol.NewAtom()
628             p4.SetAtomicNum(11)
629             p4.SetVector(float(proj_PROT[0]),float(proj_PROT[1]),float(proj_PROT[2]))
630
631             ###-----
632             ### +/- 10 degrees; opt 90 degrees
633             ### Edge 2 Face
634             if 80.0 <= alpha <= 100.0:
635                 ### less restrictive
636                 if 140.0 <= beta <= 180.0:
637                     if element not in PI_PI_interact:
638                         PI_PI_interact.append(element)
639             ### +/- 10 degrees
640             ### Face 2 Face

```

```

641         if 90.0 <= beta <= 180.0:
642             min_gamma = beta -15
643             max_gamma = beta +15
644             if min_gamma <= gamma <= max_gamma:
645
646                 if element not in PI_PI_interact:
647                     PI_PI_interact.append(element)
648
649     except:
650         print "No neighbors within search radius !"
651         print "\n"
652 #####
653 ##### DEBUGGING ONLY!
654 ##### write debug mol ...
655 pybelmol = pybel.Molecule(mol)
656 pybelmol.write("sdf", 'zzz_debug_PI_PI.sdf' , overwrite = True)
657 print "\n"
658 print "Debug PI-PI SUCCESSFULLY WRITTEN"
659 for element in PI_PI_interact:
660     pi = mol_interact.NewAtom()
661     pi.SetAtomicNum(9)
662     pi.SetVector(float(element[0]), float(element[1]), float(element[2]))
663 ##### write debug mol ...
664 pybelmol_II = pybel.Molecule(mol_interact)
665 pybelmol_II.write("sdf", 'zzz_interact_PI_PI.sdf' , overwrite = True)
666 print "\n"
667 print "PI-PI INTERACTIONS SUCCESSFULLY WRITTEN"
668 return PI_PI_interact
669
670 ##### 13-sept-2010 --> check out cationic <--> pi interactions; ache: leve as reference
671 ##### rewritten 11-may-2011 (fk)
672 ##### --> to be used for pi-pi interactions as well!
673 ##### --> adapted to pharao core; do not consider aromatic norm!
674 def sift_CAT_PI(POSC, structure, rules, RANGE=[2.5, 4.5]):
675     print "start checking cat-pi interactions ... "
676     print "\n"
677     CAT_PI_interact = []
678     test = openbabel.OBMol()
679     #PHE_local = ['CG', 'CZ', 'CD1', 'CD2', 'CE1', 'CE2']
680     #TRP_local = ['CE2', 'CD2', 'CE3', 'CZ3', 'CH2', 'CZ2']
681     #HIS_local = ['CE1', 'CD2', 'CG', 'ND1', 'NE2']
682     AROMATIC_local = ['CG', 'CZ', 'CD1', 'CD2', 'CE1', 'CE2', 'PHE / TYR --> identical!!',
683                       'CE2', 'CD2', 'CE3', 'CZ3', 'CH2', 'CZ2', 'TRP',
684                       'CE1', 'CD2', 'CG', 'ND1', 'NE2']
685     for element in POSC:
686         center = asarray([element[0], element[1], element[2]])
687         ##### Neighbour search!
688         query_atoms=Selection.unfold_entities(structure, 'A') ##### Seraching complete Structure!!
689         ns=NeighborSearch(query_atoms)
690         res_control = []
691         ##### CENTER --> UPPER DIST RANGE
692         neighbors = ns.search(center, RANGE[-1])
693         print neighbors
694         try:
695             print "\n"
696             atom_list=Selection.unfold_entities(neighbors, 'A') ##### Atom List containing ALL neighbours
697             !
698             print atom_list, "All Neighbors to center"
699             ##### Atoms which make contact --> + getting Parents // entire residue
700             for atom in atom_list:
701                 atom_id = atom.get_id()
702                 ##### getting the Parent AS from atom within radius
703                 res = atom.get_parent()
704                 chain = res.get_parent()
705                 res_name = res.get_resname()
706                 res_id = res.get_id()[1]
707                 ##### Exclusion 1: right AS Atom --> if one residue is matching; check whole res!
708                 if res_name in rules:
709                     ##### check every res only once!
710                     if res_id not in res_control:
711                         res_control.append(res_id)
712                         hit_res = chain[res_id]
713                         ##### for all ring atoms to check!!
714                         hit_list = []
715                         for hit_atom in hit_res:
716                             hit_id = hit_atom.get_id()
717                             print hit_id
718                             if hit_id in AROMATIC_local:

```

```

717         hit_coords = hit_atom.get_coord()
718         entry = [hit_coords[0], hit_coords[1], hit_coords[2]]
719         print entry, hit_id
720         hit_list.append(entry)
721     hit_array = asarray(hit_list)
722     print hit_array
723     ### coord list for hits
724     coord_list = []
725     mean_point = hit_array.mean(axis=0)
726     entry_mean = [mean_point[0], mean_point[1], mean_point[2]]
727     coord_list.append(entry_mean)
728     ### point annotation --> arbitrarily
729     CG = hit_array[0,:]
730     CG_entry = [CG[0], CG[1], CG[2]]
731     coord_list.append(CG_entry)
732     CE1 = hit_array[4,:]
733     CE1_entry = [CE1[0], CE1[1], CE1[2]]
734     coord_list.append(CE1_entry)
735     a = CG - mean_point
736     b = CE1 - mean_point
737     to_proj_a = cross(a, b)
738     to_proj_b = cross(b, a)
739     norm_a = math.sqrt(sum((to_proj_a)**2))
740     norm_b = math.sqrt(sum((to_proj_b)**2))
741     ### Normalize projected points !!!
742     ### 2.1A --> analogue to moe proj pts
743     new_a = (to_proj_a/ norm_a)*2.1
744     new_b = (to_proj_b/ norm_b)*2.1
745     ### adding to ring center
746     proj_pt_a = new_a + mean_point
747     proj_pt_b = new_b + mean_point
748     ### two points -- for both ring planes
749     proj_entry_a = [proj_pt_a[0], proj_pt_a[1], proj_pt_a[2]]
750     coord_list.append(proj_entry_a)
751     proj_entry_b = [proj_pt_b[0], proj_pt_b[1], proj_pt_b[2]]
752     coord_list.append(proj_entry_b)
753     ### check proj pt closer to feat...
754     difference_a = math.sqrt(sum((proj_pt_a - center)**2))
755     difference_b = math.sqrt(sum((proj_pt_b - center)**2))
756     if difference_a <= difference_b:
757         closer_proj_pt = proj_pt_a
758     else:
759         closer_proj_pt = proj_pt_b
760     ### calculate and check angle
761     vec_1 = closer_proj_pt - mean_point
762     vec_2 = closer_proj_pt - center
763     skalar = numpy.inner(vec_1, vec_2)
764     norm_1 = math.sqrt(sum((vec_1)**2))
765     norm_2 = math.sqrt(sum((vec_2)**2))
766     cat_pi_angle = math.degrees(math.acos((skalar)/(norm_1*norm_2)))
767     if cat_pi_angle > 180:
768         cat_pi_angle = 360 - cat_pi_angle
769     ### Angle check
770     if cat_pi_angle >= 125 and cat_pi_angle < 180:
771         min_dist = min((cdist(asarray([center]), hit_array))[0,:])
772         if min_dist >= RANGE[0]:
773             ### REMOVE SIFT REDUNDANCIES
774             if entry not in CAT_PI_interact:
775                 CAT_PI_interact.append(element)
776     ### depict center
777     x = test.NewAtom()
778     x.SetAtomicNum(13)
779     x.SetVector(float(center[0]), float(center[1]), float(center[2]))
780     ###
781     for i in coord_list:
782         print i, 'coord entry'
783         x = test.NewAtom()
784         x.SetAtomicNum(15)
785         x.SetVector(float(i[0]), float(i[1]), float(i[2]))
786     ###
787     ### no neighbours
788     except:
789         print "No neighbours 4 POS IONIZIBLE Sphere found within tolerance sphere!"
790     pybelmol = pybel.Molecule(test)
791     pybelmol.write("sdf", 'debug_cat_pi.sdf', overwrite = True)
792     print "successfully written ..."
793     print "\n"

```

```

794     ### writing approved ional features
795     for k in CAT_PI_interact:
796         x = molecule_III.NewAtom()
797         x.SetAtomicNum(OBcol)
798         x.SetVector(float(k[0]),float(k[1]),float(k[2]))
799     pybelmol = pybel.Molecule(molecule_III)
800     pybelmol.write("sdf",name_2_save , overwrite = True)
801     print "\n"
802     print "SIFTS SUCCESSFULLY WRITTEN"
803     return CAT_PI_interact
804
805     ### 08-sep-09 (fk)
806     ### recent update: additional Lig_X-Don-H angle; requires a new dict-the-lig-routine ,
807     ### to detect Acc_X atoms!!!
808     ### 16-may-2011
809     ### reset X=O norm!! (retain reasonable angles!)
810     ### 10-may-2011 (fk) // shifted norms 16-may-2011 (fk)
811     ### rewritten for cavka pharao
812     def sift_HACC(HACC, structure, ref_ligand, rules, RANGE):
813         print "start detecting H bond acceptors ...",
814         print "\n"
815         ### Debug !
816         molecule = openbabel.OBMol()
817         HACC_interact= []
818         ### 16-may-2011
819         ### reshift ALL ACCEPTORS!
820         ### working for acids , (ethers , hydroxyls--> 1di8)
821         for element in HACC:
822             center = asarray([float(element[0]), float(element[1]), float(element[2])])
823             x = molecule.NewAtom()
824             x.SetAtomicNum(12)
825             x.SetVector(float(center[0]),float(center[1]),float(center[2]))
826             ### old norm
827             norm_acc= asarray([float(element[5]), float(element[6]), float(element[7])])
828             x = molecule.NewAtom()
829             x.SetAtomicNum(8)
830             x.SetVector(float(norm_acc[0]),float(norm_acc[1]),float(norm_acc[2]))
831             nu_vec = norm_acc - center
832             nu_norm = math.sqrt(sum((nu_vec)**2))
833             nu_point = center - ((nu_vec)*nu_norm)
834             x = molecule.NewAtom()
835             x.SetAtomicNum(9)
836             x.SetVector(float(nu_point[0]),float(nu_point[1]),float(nu_point[2]))
837             ### 16-may-2011 (fk)
838             ### reshift!
839             norm_acc= nu_point
840             ### Neighbour search!
841             query_atoms=Selection.unfold_entities(structure,'A') ### Seraching complete Structure!!
842             ns=NeighborSearch(query_atoms)
843             ### Searching to upper Limit!
844             neighbors = ns.search(center,RANGE[-1]) ### upper dist range!
845             try:
846                 atom_list=Selection.unfold_entities(neighbors,'A') ### Atom List containing ALL neighbours
847                 !
848                 ### Atoms which make contact --> + getting Parents // entire residue
849                 for atom in atom_list:
850                     atom_id = atom.get_id()
851                     ###getting the Parent AS from atom within radius
852                     res = atom.get_parent()
853                     res_name = res.get_resname()
854                     res_id = res.get_id()[1]
855                     ### Matching the right AS atom
856                     match_str = str(res_name)+"_"+(atom_id)
857                     print match_str, res_id
858                     ### Exclusion 1: right AS Atom
859                     if match_str in dict_donor:
860                         ### heavy donor atom
861                         don_coords = numpy.asarray(atom.get_coord())
862                         ### vis donor
863                         x = molecule.NewAtom()
864                         x.SetAtomicNum(7)
865                         x.SetVector(float(don_coords[0]),float(don_coords[1]),float(don_coords[2]))
866                         dist = math.sqrt(sum((center-don_coords)**2))
867                         ### Exclusion 2: above clash distance
868                         if dist >= RANGE[0]:
869                             ### collecting all coords 4 angle calc ###
870                             ### H_neighbors (2 heavy donor atom) via dict

```

```

870 pot_H = dict_donor[match_str]
871 print pot_H
872 print "\n"
873 for H in pot_H:
874     ### except potential tautomeric states (caused by prot 3D)
875     ### no standard
876     try:
877         H_atom = res[H]
878         H_don_coors= numpy.asarray(H_atom.get_coord())
879         ### vis H atoms attached 2 donor
880         x = molecule.NewAtom()
881         x.SetAtomicNum(7)
882         x.SetVector(float(H_don_coors[0]),float(H_don_coors[1]),float(H_don_coors[2]))
883         ### angle calculation
884         vec_1 = don_coors - H_don_coors
885         vec_2 = center - H_don_coors
886         skalar = numpy.inner(vec_1, vec_2)
887         norm_1 = math.sqrt(sum((vec_1)**2))
888         norm_2 = math.sqrt(sum((vec_2)**2))
889         angle = math.degrees(math.acos((skalar)/(norm_1*norm_2)))
890         ### correct angle Don-H-Acc
891         if 110.0 < angle < 180.0:
892             ### 08-sep-09 (fk)
893             ### Calculate H-Acc(center)-Norm(shift) angle
894             vec_3 = H_don_coors - center
895             vec_4 = norm_acc - center
896             skalar_II = numpy.inner(vec_3, vec_4)
897             norm_3 = math.sqrt(sum((vec_3)**2))
898             norm_4 = math.sqrt(sum((vec_4)**2))
899             angle_II= math.degrees(math.acos((skalar_II)/(norm_3*norm_4)))
900             ### changed for serine proteases; 25-jan-10
901             if 90.0 < angle_II < 170.0:
902                 ### final appending
903                 if element not in HACC_interact:
904                     HACC_interact.append(element)
905                 else:
906                     print '2nd angle not passed', angle_II
907             else:
908                 print '1st angle not passed', angle
909
910         ### 08-sep-09 (fk)
911         ### exception of tautomeric states using KDTree
912         ### Not possible via dict
913         except KeyError:
914             print 'Warning tautomeric state !'
915             print "\n"
916             print "\n"
917             ### check!!!
918             ### use KDtree to search from heavy Don Center up to pot H
919             neighbors_xcept = ns.search(don_coors,1.2)
920             for Nu_H in neighbors_xcept:
921                 Nu_H_id = Nu_H.get_id()
922                 ### XCLUDE VIA String check CHECK!!!
923                 Nu_H_name = Nu_H.get_fullname()
924                 print Nu_H_name
925                 if not 'N' in Nu_H_id and not 'O' in Nu_H_id and 'H' in Nu_H_id :
926                     print "\n"
927                     print "potential H tautomerized -- not in std donor_dict", Nu_H_id
928                     print "\n"
929                 ### already array
930                 Nu_H_coors = Nu_H.get_coord()
931                 ### angle calculation ACC-H-DON
932                 vec_1 = don_coors - Nu_H_coors
933                 vec_2 = center - Nu_H_coors
934                 skalar = numpy.inner(vec_1, vec_2)
935                 norm_1 = math.sqrt(sum((vec_1)**2))
936                 norm_2 = math.sqrt(sum((vec_2)**2))
937                 angle = math.degrees(math.acos((skalar)/(norm_1*norm_2)))
938                 ### correct angle Don-H-Acc
939                 if 135.0 < angle < 180.0:
940                     ### Norm-ACC-H
941                     vec_3 = norm_acc - center
942                     vec_4 = Nu_H_coors - center
943                     skalar_II = numpy.inner(vec_3, vec_4)
944                     norm_3 = math.sqrt(sum((vec_3)**2))
945                     norm_4 = math.sqrt(sum((vec_4)**2))
946                     angle_II = math.degrees(math.acos((skalar_II)/(norm_3*norm_4)))

```

```

947         ### check H-Acc-NORM angle
948         if 90.0 < angle_II < 170.0:
949             ### final appending
950             if element not in HACC_interact:
951                 HACC_interact.append(element)
952         else:
953             print neighbors_xcept, "NO H atoms found!"
954     except:
955         print "No atoms found in tolerance radius!"
956         print "\n"
957     ### debugging POSSIBLE interactions
958     pybelmol = pybel.Molecule(molecule)
959     pybelmol.write("sdf","acc_DEBUG_RE_shift.sdf", overwrite = True)
960     print "acceptor interactions successfully written to file"
961     ### debugging TRUE interactions
962     molecule_II = openbabel.OBMol()
963     for k in HACC_interact:
964         x = molecule_II.NewAtom()
965         x.SetAtomicNum(9)
966         x.SetVector(float(k[0]),float(k[1]),float(k[2]))
967     pybelmolII = pybel.Molecule(molecule_II)
968     pybelmolII.write("sdf","acc_INTERACT.sdf", overwrite = True)
969     return HACC_interact
970 ### 10-may-2011 (fk)
971 def sift_POSC(POSC,structure, rules, RANGE):
972     print "\n"
973     print "start new sifting function ionic positive."
974     print "\n"
975     ### all cords from feat. which can interact with site
976     POSC_interact = []
977     for element in POSC:
978         center = asarray([element[0], element[1], element[2]])
979         ### Neighbour search!
980         query_atoms = Selection.unfold_entities(structure, 'A')
981         ns = NeighborSearch(query_atoms)
982         ### Searching to upper Limit!
983         neighbors = ns.search(center, RANGE[-1])    ### upper dist range!
984         print neighbors
985     try:
986         atom_list = Selection.unfold_entities(neighbors, 'A')### Atom List containing ALL neighbours
987         !
988         ### Atoms which make contact --> + getting Parents // entire residue
989         for atom in atom_list:
990             atom_id = atom.get_id()
991             ### getting the Parent AS from atom within radius
992             res = atom.get_parent()
993             res_name = res.get_resname()
994             res_id = res.get_id()[1]
995             ### Matching the right AS atom
996             match_str = str(res_name)+"_"+(atom_id)
997             if match_str in rules:
998                 atom_coords = numpy.asarray(atom.get_coord())
999                 ### clash distance
1000                 dist = math.sqrt(sum((center-atom_coords)**2))
1001                 ### Exclusion 2: above clash distance
1002                 if dist >= RANGE[0]:
1003                     ### appending 2 filtered SIFT list + removing Doubles
1004                     if element not in POSC_interact:
1005                         POSC_interact.append(element)
1006                     else:
1007                         print "Double Entry -- not appending to list"
1008         except:
1009             print "No neighbours on bindingsite found within search radius of ", RANGE[-1]
1010     return POSC_interact
1011
1012 ### 09-may-11 (fk)
1013 ### simple sifting function for IONIC negative features ;
1014 ### using distance critaria only; no angle constraints;
1015 ### currently -- considering + / - interactions only
1016 ### rewritten for cavKA pharao
1017 ### 18-may-2011(fk)
1018 ### ionic (-) -- donor interactions --> to be flagged as acceptors!
1019
1020 def sift_NEGC(NEGC,structure, rules, RANGE):
1021     print "\n"
1022     print "start new sifting function ionic negative."

```

```

1023     print "\n"
1024     metals = [ 'ZN_ZN', 'MG_MG' ]
1025     ### all cords from feat. which can interact with site
1026     NEGC_interact = []
1027     mol = openbabel.OBMol()
1028     for element in NEGC:
1029         center = asarray([element[0], element[1], element[2]])
1030         x = mol.NewAtom()
1031         x.SetAtomicNum(18)
1032         x.SetVector(center[0], center[1], center[2])
1033         ### Neighbour search!
1034         query_atoms = Selection.unfold_entities(structure, 'A')
1035         ns=NeighborSearch(query_atoms, bucket_size=10)
1036         ### Searching to upper Limit!
1037         neighbors = ns.search(center, RANGE[-1])    ### upper dist range!
1038     try:
1039         atom_list=Selection.unfold_entities(neighbors, 'A')    ### Atom List containing ALL neighbours
1040         !
1041         ### Atoms which make contact --> + getting Parents // entire residue
1042         for atom in atom_list:
1043             atom_id = atom.get_id()
1044             ### getting the Parent AS from atom within radius
1045             res = atom.get_parent()
1046             res_name = res.get_resname()
1047             res_id = res.get_id()[1]
1048             ### Matching the right AS atom
1049             match_str =str(res_name)+"_"+ (atom_id)
1050             if match_str in rules or metals:
1051                 atom_coords= numpy.asarray(atom.get_coord())
1052                 y = mol.NewAtom()
1053                 y.SetAtomicNum(12)
1054                 y.SetVector(float(atom_coords[0]), float(atom_coords[1]), float(atom_coords[2]))
1055                 ### clash distance
1056                 dist = math.sqrt(sum((center-atom_coords)**2))
1057                 ### Exclusion 2: above clash distance
1058                 if dist >= RANGE[0]:
1059                     ### appending 2 filtered SIFT list + removing Doubles
1060                     if element not in NEGC_interact:
1061                         NEGC_interact.append(element)
1062                     else:
1063                         print "Double Entry -- not appending to list"
1064                 if match_str in metals:
1065                     if dist >= 1.5:
1066                         if element not in NEGC_interact:
1067                             NEGC_interact.append(element)
1068             except:
1069                 print "No neighbours on bindingsite found within search radius of ", RANGE[-1]
1070     pybelmol = pybel.Molecule(mol)
1071     pybelmol.write("sdf", "anionic_debug.sdf", overwrite = True)
1072     mol_II = openbabel.OBMol()
1073     for i in NEGC_interact:
1074         print i
1075         x = mol_II.NewAtom()
1076         x.SetAtomicNum(18)
1077         x.SetVector(i[0], i[1], i[2])
1078     pybelmol_II = pybel.Molecule(mol_II)
1079     pybelmol_II.write("sdf", "anionic_INTERACT.sdf", overwrite = True)
1080     return NEGC_interact

```

26.2 Modul cavka_E.py

```

1  ### cavka_E.py
2  ### 20-may-2011 (fk)
3  ### module for mapping normalized and contoured MIFs to
4  ### compatible ligand-features
5
6  import openbabel
7  import pybel
8  from numpy import*
9  from hcluster import *
10
11 def read_HOTKONT(hot_kont, features, feat_type):
12     print "\n"
13     print "reading out hot kont ...", hot_kont

```



```

14 print "\n"
15 color_dict = {'AROM':9, 'LIPO':6, 'HACC':8, 'HDON':7, 'POSC':35, 'NEGC':15, 'EXCL':1}
16 ### NU LABELED LIST TO BE RETURNED ...
17 E_kont = []
18 ### debug neighbour_grid_pts!
19 mol = openbabel.OBMol()
20 kont_list = []
21 kont_file = file('hot_kont', 'r')
22 kont_data = kont_file.readlines()
23 for line in kont_data:
24     i = line.split(' ', ' ')
25     entry = [float(i[0]), float(i[1]), float(i[2]), float(i[3])]
26     kont_list.append(entry)
27 kont_list.sort()
28 kont_list.reverse()
29 ### 03-feb-2012 (fk)
30 ### normalization done in prepare_cavka_grid.py
31 kont_array = asarray(kont_list)
32 kont_coords = kont_array[:,1:]
33 ### match 2 field coords and scale
34 feat_list = []
35 for element in features:
36     print element
37     feat_coords = [element[0], element[1], element[2]]
38     feat_list.append(feat_coords)
39 feat_array = asarray(feat_list)
40 dist_array = cdist(feat_array, kont_coords)
41 for i in range(len(feat_list)):
42     cur_row = dist_array[i]
43     test = argwhere(cur_row <= 1.5)
44     ### test --> minimum of surrounding grid points
45     ### (currently no limit of patch size)
46     if len(test) > 0:
47         test_nu = (test[:,0]).T
48         ### re-indexing 2 energy
49         neighbour_grid_pts = (kont_array[test_nu])
50
51     for t in neighbour_grid_pts:
52         color = color_dict[feat_type]
53         x = mol.NewAtom()
54         x.SetAtomicNum(color)
55         x.SetVector(float(t[1]), float(t[2]), float(t[3]))
56     E_label = ([neighbour_grid_pts.mean(axis=0)][0])[0]
57     cur_e = features[i]
58     ### re-organize
59     cur_E_kont = [E_label,
60     cur_e[0], cur_e[1], cur_e[2], ### coords
61     cur_e[3], cur_e[4], ### alpha/ norm(0/1)
62     cur_e[5], cur_e[6], cur_e[7], ### normal coords
63     feat_type]
64     E_kont.append(cur_E_kont)
65 ### Debug
66 name_2_save = 'xxx_E_'+feat_type+'_DEBUG.sdf'
67 pybelmol = pybel.Molecule(mol)
68 pybelmol.write("sdf", name_2_save, overwrite = True)
69 print name_2_save, 'successfully written ...'
70 print "\n"
71 return E_kont

```

26.3 Programm prepare_grid_cavka.py

```

1 ### prepare_cavka_grid.py
2 ### 09-may-2011 (fk)
3 ### 28-jan-2012 (fk) (MODIFIED)
4
5 import numpy
6 import openbabel
7 import pybel
8 import sys, os, time
9 from numpy import *
10 from time import *
11
12 ### for cleaning structure
13 import Bio.PDB

```

```

14 from Bio.PDB import*
15 from Bio.PDB.StructureBuilder import *
16 from hcluster import*
17 from operator import itemgetter
18 ### 1) cleaning
19 ### 2) GRID calculations
20 ### 3) Mapping
21 def clean_pdb(pdb_2_clean, mol_ligand):
22     print pdb_2_clean
23     print mol_ligand
24     #### searching from feature points instead of geometrical centers!
25     #### No filtering for regular AS Residues anymore -->reimplement!
26     requirement_list = [ 'ARG', 'R', 'GLN', 'Q', 'PHE', 'F', 'TYR', 'Y', 'TRP', 'W', 'LYS', 'K', 'GLY', 'G', 'ALA',
        'A', 'HIS', 'H', 'SER', 'S', 'PRO', 'P', 'GLU', 'E', 'ASP', 'D', 'THR', 'T', 'CYS', 'C', 'MET', 'M',
        'LEU', 'L', 'ASN', 'N', 'ILE', 'I', 'VAL', 'V' ]
27
28     ligand_coords = []
29     lig_data = pybel.readfile('sdf',mol_ligand)
30     for mol in lig_data:
31         for atom in mol:
32             coords = atom.coords
33             x = coords[0]
34             y = coords[1]
35             z = coords[2]
36             entry = [x,y,z]
37             ligand_coords.append(entry)
38     ligand_array = numpy.asarray(ligand_coords)
39     ###structure object
40     parser =PDBParser(PERMISSIVE=1)
41     structure = parser.get_structure("s",pdb_2_clean)
42     print "\n"
43     print " protein successfully read in ! "
44     print "\n"
45     ### cleaned output
46     part1 = pdb_2_clean.replace(".pdb", "")
47     clean_output_protein = part1+"_clean"+" .pdb"
48     ### neighbour search--> complete structure
49     query_atoms=Selection.unfold_entities(structure, 'A')
50     ns=NeighborSearch(query_atoms)
51
52     #####
53     ### Checking for possible chains ###
54     ###searching from EVERY lig atom up to 5A ###
55     #####
56
57     ### "binding site chains"
58     chain_list_I = []
59     for center in ligand_array:
60         try:
61             ####5A tol --> check out ludi paper!!
62             neighbors = ns.search(center,5.0)
63             ####GET chain ID from neighbours!!!
64             pre_chain_list=Selection.unfold_entities(neighbors, 'C')
65             for chain in pre_chain_list:
66                 chain_id = chain.get_id()
67                 ### excepting hetero flags (=ligand chain)
68                 if chain_id != " ":
69                     ####avoid having doubles
70                     if chain_id not in chain_list_I:
71                         chain_list_I.append(chain_id)
72         except:
73             print "No neighbours matched within search radius"
74     chain_list_I.sort()
75     print chain_list_I , " Chains forming the active site "
76     print "\n"
77     ### Slicing over atoms for removing potential HET Entries within
78     ### all for build up--> chain by chain!
79     io =Bio.PDB.PDBIO()
80     m = Model(0)
81     s = Structure(clean_output_protein)
82     s.add(m)
83     ### make sure that ligand is surrounded with atoms!
84     if len(chain_list_I)>=1:
85         ### Getting ALL CHAINS from structure
86         all_chains=Selection.unfold_entities(structure, 'C')
87         for element in all_chains:
88             ### getting chain id --> and comparing 2 list

```

```

89     entry_2_check = element.get_id()
90     if entry_2_check in chain_list_I:
91         ### No strings for chain description!! Use int or char
92         new_id = element.get_id()
93         name_for_new_chain = new_id
94         new_chain = Chain(name_for_new_chain)
95         ### atom level—> Atoms in each chain
96         for atoms in element:
97             atom_id = atoms.get_id()
98             ### APPENDING here!
99             ### getting the NON- Hets—> if Hets- Non Hets are mixed!
100             if atom_id[0] == " ":
101                 new_chain.add(atoms)
102             m.add(new_chain)
103         io.set_structure(s)
104         io.save(clean_output_protein)
105         print clean_output_protein, "written and saved -> cleaned protein"
106         return clean_output_protein
107     else:
108         print "ERROR! No chain near ligand!"
109         error_string_repairing = file_2_kont+"No chain near ligand !!"
110         datei.write(error_string_repairing)
111         assert()
112
113 ### gridding
114 ### modified; 25-nov-09 (fk)
115 def box_and_features(ligand, tolerance):
116     mol_content = pybel.readfile("mol", ligand)
117     ligand_coords = []
118     for molecule in mol_content:
119         for atom in molecule:
120             ligand_coords.append(atom.coords)
121     ### create array
122     ligand_array = numpy.asarray(ligand_coords)
123     min_val = ligand_array.min(axis=0) - tolerance
124     max_val = ligand_array.max(axis=0) + tolerance
125     ### changed! 09-may-2011 (fk)
126     ### take care of dimensionality for ascii grid
127     box = [round(min_val[0]), round(min_val[1]), round(min_val[2]), round(max_val[0]), round(
128         max_val[1]), round(max_val[2])]
129
130     ### writing box data 2 file
131     box_data = file("box_data.txt", "w")
132     box_str = str(ligand) + "    min    " + str(box[0]) + "    " + str(box[1]) + "    " + str(box[2]) + "    max"
133     box_data.write(box_str)
134     for i in range(2):
135         box_data.write("\n")
136     box_data.close()
137     return box, ligand_array
138
139 ### completely revised; 30-nov-09 (jd) / (fk)
140 ### integrated again 11-may-11(fk)
141 def generate_kontfile_NEW(_GRIN, _GRID, instance, indir, outdir, binary, cur_probe):
142     ##Attention check the PATH!
143     print "\n"
144     print "started Grid run ... "
145     print "\n"
146     print _GRIN
147     print _GRID
148     print "\n"
149     file_grin = open("grin.in", "w")
150     file_grin.write(_GRIN)
151     file_grin.close()
152     #### PATH
153     os.system("/opt/grid/grin < grin.in")
154     #### FILE LIST —> contains input for grid runs [.kout files]
155     inpt = "file.list"
156     file_list = open(inpt, "w")
157     kout_str = cur_name + ".kout"
158     file_list.write(kout_str)
159     file_list.close()
160     file_grid = open("grid.in", "w")
161     ### Has to be set to -2 when moe kontfile viewer is used!!
162     ### (binary files required)
163     if binary == 0:

```

```

164     LIST = -2
165 else:
166     LIST = 1
167     file_grid.write(_GRID)
168     file_grid.close()
169     print "\n"
170     print cur_probe, " CURRENT KONT GENERATION .... "
171     print "\n"
172     os.system("/opt/grid/grid < grid.in")
173
174 def read_kont(C3_kont):
175     ### Legend
176     ### [0]    --> Energy
177     ### [1,2,3] --> x/y/z
178     print "\n"
179     print "read out C3 kont "
180     print "\n"
181     file_2_read = C3_kont
182     COORD_list = []
183     IDX_list = []
184     FIELD_list = []
185     file_obj = file(file_2_read, 'r')
186     lines = file_obj.readlines()
187     for l in lines:
188         l_split = l.split()
189         if len(l_split) == 4 and l_split[1] != 'HEADER':
190             ##### coords
191             coord_entry = [float(l_split[1]), float(l_split[2]), float(l_split[3])]
192             COORD_list.append(coord_entry)
193             ##### indices
194             idx_entry = [int(l_split[0])]
195             IDX_list.append(idx_entry)
196             if len(l_split) == 1:
197                 coord_entry = [float(l_split[0])]
198                 FIELD_list.append(coord_entry)
199     print "FILE SUCCESULLY READ OUT."
200     print "\n"
201     print len(COORD_list), 'coord'
202     print "\n"
203     print len(FIELD_list), 'field'
204     if len(COORD_list) == len(FIELD_list):
205         coord_array = asarray(COORD_list)
206         field_array = asarray(FIELD_list)
207         grid_idx_array = asarray(IDX_list)
208         kont_array = hstack((field_array, coord_array, grid_idx_array))
209         print shape(kont_array), 'before zero elimination'
210     else:
211         print "\n"
212         print ".kont composition ERROR!!!!"
213         print "\n"
214         assert()
215     ### keep all C3 points; E<5 moon value (no energy adaption)!
216     cond_zero = (kont_array[:,0] <= 5)
217     kont_array = kont_array[cond_zero]
218     print "\n"
219     print "kont array successfully read out"
220     print "\n"
221     return kont_array
222
223 def map_act_site_to_C3(cav_array, kont_array, mapping_dist):
224     check_idx = asarray([zeros([shape(kont_array)[0]])])
225     kont_coords = kont_array[:,1:4]
226     dist_array = cdist(kont_coords, cav_array)
227     for i in range(shape(dist_array)[0]):
228         ### important!!! mapping done here!!!
229         test = argwhere(dist_array[i,:] <= mapping_dist)
230         if len(test) > 0:
231             check_idx[0,i] = 1
232     check_idx = check_idx.T
233     new_kont_array = hstack((kont_array, check_idx))
234     cond = (new_kont_array[:, -1] == 1)
235     new_kont_array = new_kont_array[cond]
236     new_kont_array = new_kont_array[:,0:5]
237     ### Debug mapped grid
238     mol = openbabel.OBMol()
239     for i in new_kont_array:
240         x = mol.NewAtom()

```

```

241     x.SetAtomicNum(11)
242     x.SetVector(i[1], i[2], i[3])
243     pybelmol = pybel.Molecule(mol)
244     pybelmol.write("mol2", "zzz_act_site_mapped_grid.mol2" , overwrite = True)
245     print "\n"
246     print "MAPPED GRID SUCCESSFULLY WRITTEN!"
247     print "\n"
248     ### create output .txt
249     output = file('active_site.txt', 'w')
250     for i in new_kont_array:
251         output_str = str(int(i[-1]))
252         output.write(output_str)
253         output.write("\n")
254     print "\n"
255     print "OUTPUTFILE SUCCESSFULLY WRITTEN!"
256     print "\n"
257
258     #### 2nd stage --> reading out !!!
259     #### read out act site info only once !
260     def get_active_site(active_site):
261         print "\n"
262         print "starting kont cleaning routine..."
263         print "\n"
264         ### all grid points belonging to active site
265         act_site_idx = []
266         act_site_obj = file(active_site)
267         act_lines = act_site_obj.xreadlines()
268         for k in act_lines:
269             k_split = k.split()
270             if len(k_split) == 1:
271                 act_idx_entry = int(k_split[0])
272                 act_site_idx.append(act_idx_entry)
273         act_site_idx = asarray(act_site_idx).T
274         return act_site_idx
275
276     def active_site_kont(kont, act_site_idx, E_cut, color, flag):
277         ### Legend
278         ### [0] --> Energy
279         ### [1,2,3] --> x/y/z
280         ### [4] --> grid_idx array
281         ### I) adapt to bindingsite
282         ### II) scale
283         ### III) contour
284         print "\n"
285         print "reduce kont to cavity related kont ...", kont
286         print "\n"
287         print "READING IN KONT FILE ..."
288         print "\n"
289         ### read-out kont info
290         file_2_read = kont
291         COORD_list = []
292         IDX_list = []
293         FIELD_list = []
294         file_obj = file(file_2_read, 'r')
295         lines = file_obj.xreadlines()
296         for l in lines:
297             l_split = l.split()
298
299             if len(l_split) == 4 and l_split[1] != 'HEADER':
300                 #### coords
301                 coord_entry = [float(l_split[1]), float(l_split[2]), float(l_split[3])]
302                 COORD_list.append(coord_entry)
303                 #### indices
304                 idx_entry = [int(l_split[0])]
305                 IDX_list.append(idx_entry)
306
307             if len(l_split) == 1:
308                 coord_entry = [float(l_split[0])]
309                 FIELD_list.append(coord_entry)
310
311         print "FILE SUCCESULLY READ OUT."
312         print "\n"
313         if len(COORD_list) == len(FIELD_list):
314             coord_array = asarray(COORD_list)
315             field_array = asarray(FIELD_list)
316             grid_idx_array = asarray(IDX_list)
317             kont_array = hstack((field_array, coord_array, grid_idx_array))

```

```

318 ### Water probe: no adaption to bindingsite
319 if 'OH2.kont' in kont:
320     red_kont = kont_array
321 else:
322     ### I) adaption to bindingsite
323     red_kont = kont_array[act_site_idx,:]
324     ### check
325     print "\n"
326     print shape(act_site_idx)[0], "points belonging to active site "
327     print "\n"
328     if shape(red_kont)[0] == shape(act_site_idx)[0]:
329         print "\n"
330         print "LENGHT OKAY -- CONTINUE ..."
331         print "\n"
332     else:
333         print "\n"
334         print "ASSEMBLY ERROR -- STOP!"
335         print "\n"
336         assert()
337 kont_list = []
338 ### create kont list for sorting and scaling...
339 for i in red_kont:
340     if i[0]<0:
341         kont_entry = [i[0],i[1],i[2], i[3]]
342         kont_list.append(kont_entry)
343 sorted_kont_list = sorted(kont_list, key=itemgetter(0))
344
345     ### II) scaling
346     ### Dreher's scaling --> take 2% hottest point and set to one;
347     ### scale rest according to 'coldest' hotspot
348     scale_idx = int(round(len(sorted_kont_list) *0.02))
349     print scale_idx, 'scale_idx'
350
351     scale_factor = (sorted_kont_list[scale_idx])[0]
352     print scale_factor, 'scale_factor'
353
354     scaled_kont = []
355     counter = 0
356     for j in sorted_kont_list:
357         if counter<= scale_idx:
358             j[0] = 1
359         else:
360             j[0] = j[0]/scale_factor
361             entry = [j[0],j[1],j[2],j[3]]
362             counter = counter+1
363     ### III) countouring
364     if entry[0] >= E_cut:
365         scaled_kont.append(entry)
366 scaled_kont_array = asarray(scaled_kont)
367 kont_coords = scaled_kont_array[:,1:]
368 ### Visualize scaled and contoured MIFs
369 mol = openbabel.OBMol()
370 for point in scaled_kont_array:
371     x = float(point[1])
372     y = float(point[2])
373     z = float(point[3])
374
375     t = mol.NewAtom()
376     t.SetAtomicNum(color)
377     t.SetVector(x, y, z)
378
379 name_2_write = 'zzz_'+flag+'_HOT_kont.mol2'
380 pybelmol = pybel.Molecule(mol)
381 pybelmol.write("mol2",name_2_write, overwrite = True)
382 print "\n"
383 print "VISUALIZATION SUCCESSFULLY WRITTEN"
384 print "\n"
385 print "writing contoured MIF to output file ..."
386 print "\n"
387 ### new: export clean kont as .txt file for CavKA ph4 method
388 name_output_txt = kont.replace('.kont', '_HOT_kont.txt')
389 output_txt = file(name_output_txt, 'w')
390 ### remove grid idx ! [0] Energy; [1,2,3] coords
391 for i in scaled_kont_array:
392     output_str = str(i[0])+', '+str(i[1])+', '+str(i[2])+', '+str(i[3])
393     output_txt.write(output_str)
394     output_txt.write("\n")

```

```

395     print "\n"
396     print "all countoured MIF information successfully written to file ..."
397     print "\n"
398     output_txt.close()
399     return scaled_kont_array
400
401     ### DEFINE PATH VARIABLES AND DIRECTIVES
402     ### conversion
403     babel = "/opt/openeye/babel_3_3/bin/babel "
404     ### GRID CALCULATIONS
405     probe_dict = ['DRY', 'C3', 'N1', 'N1+', 'O', 'O-', 'OH2' ]
406     ### resolution == NLPA!; 4 == 0.25A GRID SPACING; 2 == 0.5A GRID SPACING
407     resolution = 2
408     binary = 1
409     ### box around ligand
410     tolerance = 5.0
411     ### mapping dist
412     mapping_dist = 1.5
413
414     ### MAIN PROGRAMM
415     start_dir = os.getcwd()
416     all_files = os.listdir(start_dir)
417     relevant_folders = [name for name in all_files if not '.py' in name and not '~' in name and
418         not '.pyc' in name and not '.txt' in name]
419     relevant_folders.sort()
420
421     for folder in relevant_folders:
422         os.chdir(folder)
423         sub_dir = os.getcwd()
424         dive_entries = os.listdir(sub_dir)
425         ### changed for org pdb complexes (--> chek Dissertation)
426         ### protonate 3D (execution and inspection: has to be done manually)
427         ### protein without H (for grid)
428         refined_LIGAND_3D = ([name for name in dive_entries if "refined_LIGAND_3D.mol" in name])[0]
429         refined_PROTEIN_3D = ([name for name in dive_entries if "refined_PROTEIN_3D.pdb" in name])[0]
430         pdb_out_NO_H = refined_PROTEIN_3D.replace("refined_PROTEIN_3D.pdb", "protein_NO_H.pdb")
431
432         print refined_LIGAND_3D, " -->LIGAND"
433         print refined_PROTEIN_3D, "-->PROTEIN"
434
435         conversion_call_II = babel+" -in "+ refined_PROTEIN_3D+" -out "+pdb_out_NO_H+ " -h delete"
436         os.system(conversion_call_II)
437         print "stripped of H from protein ... "
438         print "\n"
439
440         ### cleaning (consistent files for grid required!)
441         sub_dir = os.getcwd()
442         dive_entries = os.listdir(sub_dir)
443         pdb_2_clean = ([name for name in dive_entries if 'protein_NO_H.pdb' in name])[0]
444
445         ### repairing the protein
446         try:
447             repaired_protein = clean_pdb(pdb_2_clean, refined_LIGAND_3D)
448             print "repaired_protein"
449             print "\n"
450             print "PROTEIN CLEANED"
451         except:
452             error_string_repairing = file_2_handle+"Error! Cleaning of pdb_file failed !!"
453             datei.write(error_string_repairing)
454
455         ### GRIDDING
456         ### 11-may-2011 (fk)
457         ### Define Box
458         file_contents = box_and_features(refined_LIGAND_3D, tolerance )
459         box = file_contents[0]
460         ligand_coords = file_contents[1]
461
462         ### Grid Run
463         indir = './'
464         outdir = './'
465
466         ### 25-nov-09 (fk)
467         ### ADDITIONAL(S) : grid needs a pdb structure without(!!!) H atoms --
468         ### or some serious errors will occur! (check if files are available;
469         ### create if not existing via babel )
470         ### update file contents
471         dive_entries = os.listdir(sub_dir)
472         repaired_protein = [name for name in dive_entries if '_protein_NO_H_clean.pdb' in name][0]
473         instance = repaired_protein.replace(".pdb", "")

```

```

471 print "\n"
472 print instance, "4 generating_kont"
473 print "\n"
474 name = instance.replace("_protein_NO_H_clean", "")
475 #####
476 ### EXECUTE GRID RUN ###
477 #####
478 for probe in probe_dict:
479     cur_name = name+'_'+probe
480     print cur_name, "--> current probe"
481     print "\n"
482     #####
483     ### GRID directives ###
484     #####
485     _GRIN = " LOUT      6\n LOUT  %s.lout\n KOUT      20\n KOUT  %s.kout\n INKO      11\n INKO  %s.
         pdb\n INAT      10\n INAT  /opt/grid/grub.dat\n ALHY      0.040\n NEHY      1\n IHVA
         0\n LEVL      3\n MOVE      1\n QQHY      0.000\n VDHY      0.600\n IEND"
486     ### ENTER INFO INTO FORMATTED _GRIN
487     _GRIN = (_GRIN % (cur_name, cur_name, instance))
488     _GRID = " LONT      6\n LONT  %s.lont\n KONT      20\n KONT  %s.kont\n INPT      10\n INPT
         file.list\n CLER      5.000\n DEEP      5.000\n DPRO      4.000\n DWAT      80.000\n EACH
         5.000\n EMAX      5.000\n FARH      5.000\n FARR      8.000\n ALMD      0.000\n
         KWIK      1\n LEAU      0\n LENG      12\n LEVL      1\n LIST      %d\n MOVE      0\n
         NETA      0\n NPLA      %i\n NUMB      1\n VALU      0.000\n TOPX      %2.3f\n TOPY
         %2.3f\n TOPZ      %2.3f\n BOTX      %2.3f\n BOTY      %2.3f\n BOTZ      %2.3f\n %s \n IEND\
         nTEST\n 0 1\n"
489     ### ENTER INFO INTO FORMATTED _GRID
490     if binary == 0:
491         ### 4 displaying in moe
492         _GRID = (_GRID % (cur_name, cur_name, -2, resolution, box[3], box[4], box[5], box[0],
         box[1], box[2], probe))
493     if binary == 1:
494         _GRID = (_GRID % (cur_name, cur_name, 1, resolution, box[3], box[4], box[5], box[0],
         box[1], box[2], probe))
495     ### new routine
496     kont_file = generate_kontfile_NEW(_GRIN, _GRID, instance, indir, outdir, binary, cur_name
         )
497     ### map GRID 2 FILE
498     ### all files in subdir
499     dive_entries = os.listdir(subdir)
500     ### kont information
501     C3_kont = [name for name in dive_entries if '_C3.kont' in name][0]
502     kont_array = read_kont(C3_kont)
503     ### GRID points surrounding ligand
504     act_site_idx = map_act_site_to_C3(ligand_coords, kont_array, mapping_dist)
505     ### 15-aug- 2010
506     ### 2nd stage --> clean the kont! / Based on org cav_KA.py routine
507     ### energy countours must be assigned here !
508     ### all belonging to active site
509     active_site = 'active_site.txt'
510     act_site_idx = get_active_site('active_site.txt')
511     ### update again ...
512     dive_entries = os.listdir(subdir)
513     ### Process MIFs
514     ### DRY
515     DRY_kont = ([name for name in dive_entries if '_DRY.kont' in name])[0]
516     reduced_DRY = active_site_kont(DRY_kont, act_site_idx, E_cut = 0.75, color = 9, flag = 'DRY')
517     ### C3
518     C3_kont = ([name for name in dive_entries if '_C3.kont' in name])[0]
519     reduced_C3 = active_site_kont(C3_kont, act_site_idx, E_cut = 0.85, color = 6, flag = 'C3')
520     ### DONORS
521     N1_kont = ([name for name in dive_entries if '_N1.kont' in name])[0]
522     reduced_N1 = active_site_kont(N1_kont, act_site_idx, E_cut = 0.85, color = 7, flag = 'N1')
523     ### ACCEPTORS
524     O_kont = ([name for name in dive_entries if '_O.kont' in name])[0]
525     reduced_O = active_site_kont(O_kont, act_site_idx, E_cut = 0.85, color = 8, flag = 'O')
526     ### IONIC (+)
527     N_plus_kont = ([name for name in dive_entries if '_N1+.kont' in name])[0]
528     reduced_N_plus = active_site_kont(N_plus_kont, act_site_idx, E_cut = 0.90, color = 35, flag =
         'N+')
529     ### IONIC (-)
530     O_minus_kont = ([name for name in dive_entries if '_O-.kont' in name])[0]
531     reduced_O_minus = active_site_kont(O_minus_kont, act_site_idx, E_cut = 0.85, color = 15, flag =
         'O-')
532     ### WATER (only contouring; no adaption to active site!)
533     OH2_kont = ([name for name in dive_entries if '_OH2.kont' in name])[0]
534     reduced_OH2 = active_site_kont(OH2_kont, act_site_idx, E_cut = 0.95, color = 23, flag = 'OH2')

```



```

535
536 os.chdir(start_dir)

```

26.4 Programm cavka.py

```

1  ### cavka.py
2  ### 08-may-2011(fk)
3  ### reimplemented cavKA routine
4  ### using the silicos NV PHARAO core for feature definitions
5  ### create ligbased, structure based, energy based ph4!
6
7  ### 15-may-2011 (fk)
8  ### Debugging Don-H --Acc Interactions; evetually noise through normals
9  ### 15-may-2011(fk)
10 ### to do:
11 ### - two norm for aromatic rings
12 ### - changing norm for acceptors -- done -- (re-shift)
13 ### - changing norm for donors (heavy-donor - H) instead of norm
14 ### 18-may-2011 (fk) -- Debugging NH2 Groups
15 ### IMPORTANT -- ALTHOUGH NORMS ARE ALTERED, ALTERED NORMS are NOT written to .phar
16 ### comparison DB -- .phar would not be equal anymore (due to different norm annotations!)
17 ### 18-may-2011 (fk)
18 ### BEWARE OF FAULTY refined_PROTEIN_3D.pdb files !!!
19 ### implementing : cationic -- acceptors
20 ### 20-may-2011(fk)
21 ### implenting energetic pharmacophores:
22 ### -check grid potential at feature
23 ### -norm energy according to 2 percent hottest points
24 ### -keep the best energy pharmacophores and use them for final hypothesis
25 ### -using C3 kont for LIPO and AROM !!
26 ### 11-sep-2011(fk)
27 ### -implementing PI -- PI interactions (edge 2 face; face 2 face)
28 ### 03-feb-2012 (fk)
29 ### rewrite of energetic routine (scaling done in prepare_cavka_grid.py)
30 ### 04-feb-2012 (fk)
31 ### exporting energy-based CCG .ph4 files
32
33 import openbabel
34 import pybel
35 import os,sys,time
36 from numpy import*
37 ### own modules
38 from sift_cavka_pharao_25may11 import*
39 from cavka_E import*
40
41 def create_ph4(ref_ligand, target):
42     print "encode ligand to pharmacophore ... ", ref_ligand
43     print "\n"
44     ph4 = '1'+str(target)+'_ligbased.phar'
45     ### create HYBRID features --> HYBL dont have AROMATIC norm!!!
46     ### 10-may-2011 (fk)
47     ph4_call = pharao+' -d '+ref_ligand+' -p '+ph4+' --noHybrid '
48     print ph4_call
49     print "\n"
50     os.system(ph4_call)
51     ### parsed ph4 info
52     aromatic = []
53     donor = []
54     acceptor = []
55     pos_charge = []
56     neg_charge = []
57     ### read out ligand based ph4 information
58     ### Description          spread Normal Hybrid
59     ### NORMALIZED !! --> LIST OF LISTS; Points and directional vector
60     ### AROM Aromatic ring          0.7 Yes No
61     ### HDON Hydrogen bond donor          1.0 Yes No
62     ### HACC--> Hydrogen bond acceptor 1.0 Yes No
63     ### --> DISCARDED: HYBH Hydrogen bond donor and acceptor 1.0 Yes Yes
64     AROM = []
65     HDON = []
66     HACC = []
67     ### NOT NORMALIZED!!
68     ### LIPO Lipophilic region 0.7 No No
69     ### POSC Positive charge center 1.0 No No

```

```

70  ### NEGC Negative charge center      1.0  No  No
71  ### --> DISCARDED:  HYBL Aromatic and lipophilic      0.7  No  Yes
72  ### EXCL Exclusion sphere (LATER ON ...)  1.7  No  No
73  LIPO = []
74  POSC = []
75  NEGC = []
76  ### PARSE
77  ph4_file = file(ph4, 'r')
78  ph4_data = ph4_file.readlines()
79  for line in ph4_data:
80      i = line.split("\t")
81      ### skip header
82      if len(i) == 9:
83          ### detach feature type
84          num_core= [float(i[1]),float(i[2]),float(i[3]),float(i[4]),float(i[5]),float(i[6]), float(i
                        [7]),float(i[8])]
85          if i[0] == 'AROM':
86              AROM.append(num_core)
87          if i[0] == 'HDON':
88              HDON.append(num_core)
89          if i[0] == 'HACC':
90              HACC.append(num_core)
91          if i[0] == 'LIPO':
92              LIPO.append(num_core)
93          if i[0] == 'POSC':
94              POSC.append(num_core)
95          if i[0] == 'NEGC':
96              NEGC.append(num_core)
97      return AROM,HDON, HACC, LIPO, POSC, NEGC
98
99  ### write structure_based ph4!
100 def write_SB_ph4(SB_ph4, list_2_write, prefix, write):
101     print "writing structure based ph4 ... "
102     print "\n"
103     print prefix
104     tab = "\t"
105     norm_list =['AROM','HDON','HACC']
106     color_dict = {'AROM':9, 'LIPO':6, 'HACC':8, 'HDON':7, 'POSC':35, 'NEGC':15, 'EXCL':1}
107     color = color_dict[prefix]
108     for i in list_2_write:
109         ph4_str = prefix+tab+str(i[0])+tab+str(i[1])+tab+str(i[2])+tab+str(i[3])+tab +str(i[4])+tab+
                        str(i[5])+tab+str(i[6])+tab+str(i[7])
110         print ph4_str
111         SB_ph4.write(ph4_str)
112         SB_ph4.write("\n")
113         if write == True:
114             x = mol.NewAtom()
115             x.SetAtomicNum(color)
116             x.SetVector(float(i[0]),float(i[1]),float(i[2]))
117             if prefix in norm_list:
118                 y = mol.NewAtom()
119                 y.SetAtomicNum(12)
120                 y.SetVector(float(i[5]),float(i[6]),float(i[7]))
121
122  ### 23-may-2011 (fk)
123  ### write energetic .phar
124  ### removed!
125  ### 04-feb-2012 (fk)
126
127 def write_ligbased_EX(SB_ph4_III, ligbased_phar, EXCL):
128     print "\n"
129     print "adding EXCL TO phar ... "
130     print "\n"
131     ### PARSE
132     ph4_file = file(ligbased_phar, 'r')
133     ph4_data = ph4_file.readlines()
134     for line in ph4_data:
135         i = line.split("\t")
136         ### skip header / end
137         if len(i) == 9:
138             ### write as-it-is
139             SB_ph4_III.write(line)
140     tab = "\t"
141     for i in EXCL:
142         ph4_str = 'EXCL'+tab+str(i[0])+tab+str(i[1])+tab+str(i[2])+tab+str(i[3])+tab+str(i[4])+tab +
                        str(i[5])+tab+str(i[6])+tab+str(i[7])
143         SB_ph4_III.write(ph4_str)

```

```

144     SB_ph4_III.write("\n")
145     print "\n"
146     print "ligand based ph4 + EXCL successfully written ... "
147     print "\n"
148
149     ### 14-may-2012 (fk)
150     def write_CCG_ph4(CCG_interact, EXCL, PH4, flag):
151         print "\n"
152         print "start writing CCG ph4...", target
153         print "\n"
154         print len(CCG_interact), 'number of interacting features to write ...'
155         print "\n"
156         ### debug energetic
157         color_dict = {'AROM':9, 'LIPO':6, 'HACC':8, 'HDON':7, 'POSC':35, 'NEGC':15, 'EXCL':1}
158         new_CCG_interact=[]
159         if flag == 'E':
160             E_info = file('SB_CCG_ph4_ENERGETIC.txt', 'w')
161             mol = openbabel.OBMol()
162             for i in CCG_interact:
163                 if i[0] > 0.95:
164                     color = color_dict[i[-1]]
165                     y = mol.NewAtom()
166                     y.SetAtomicNum(color)
167                     y.SetVector(float(i[1]), float(i[2]), float(i[3]))
168                     ### write
169                     pybelmol = pybel.Molecule(mol)
170                     pybelmol.write("sdf", 'zzz_E_FEATURES.sdf', overwrite = True)
171
172                     ### remove normalized MIF value for ph4 writing
173                     new_ph4_entry = i[1:]
174                     new_CCG_interact.append(new_ph4_entry)
175                     for j in i:
176                         my_str = str(j)+"\t"
177                         E_info.write(my_str)
178                     E_info.write("\n")
179             E_info.close()
180             CCG_interact = new_CCG_interact
181
182         ### map pharao features 2 CCG
183         CCG_dict = {'LIPO': ['Hyd', '66106', '1.4'], 'AROM': ['Aro', 'df20d', '1.4'],
184                    'HACC': ['Acc', 'df2f2', '1.0'], 'HDON': ['Don', 'f20df2', '1.0'],
185                    'POSC': ['Cat', 'f20df2', '1.3'], 'NEGC': ['Ani', 'df2f2', '1.3']}
186
187         ### nr of DIFFERENT feat
188         diff_feat = []
189         ### HEADER
190         a = '#moe:ph4que 2010.1'
191         ### don't know why Tag must be set to 5 (PPCD specific???)
192         b = '#pharmacophore 5 tag t value *'
193         c = 'scheme t PCHD matchsize i 0 title t $ comment s $ smask i 2047'
194         d = '#feature '+str(len(CCG_interact))+ ' expr tt color ix x r y r z r r ebits ix gbits ix'
195         PH4.write(a)
196         PH4.write(b)
197         PH4.write(c)
198         PH4.write(d)
199         PH4.write("\n")
200         PH4.write("\n")
201         PH4.write("\n")
202         for i in range(len(CCG_interact)):
203             feat_old = (CCG_interact[i])[-1]
204             feat_new = (CCG_dict[feat_old])[0]
205             nr = (CCG_dict[feat_old])[1]
206             radius = (CCG_dict[feat_old])[2]
207             terminator = ' 0 300'
208             x = (CCG_interact[i])[0]
209             y = (CCG_interact[i])[1]
210             z = (CCG_interact[i])[2]
211             ### to be written only once!
212             if i == 0:
213                 CCG_entry = feat_new + ' '+nr+ ' '+str(x)+ ' '+str(y)+ ' '+str(z)+ ' '+radius+ terminator+ ' '
214                 PH4.write(CCG_entry)
215             if i > 0:
216                 CCG_complete = feat_new + ' '+nr
217                 PH4.write(CCG_complete)
218                 PH4.write("\n")
219                 if i < len(CCG_interact):
220                     CCG_rest = str(x)+ ' '+str(y)+ ' '+str(z)+ ' '+radius+ terminator+ ' '

```

```

221         if i == len(CCG_interact)-1:
222             CCG_rest = str(x)+' '+str(y)+' '+str(z)+' '+radius+ terminator
223             PH4.write(CCG_rest)
224         PH4.write("\n")
225         ### 15-may-2011
226         ### excluded volumes
227         ### default radius: 1A
228         EXCL_nr = len(EXCL)
229         e = '#volumesphere '+ str(EXCL_nr)+' x r y r z r r r'
230         PH4.write(e)
231         PH4.write("\n")
232         for i in EXCL:
233             cur_str = str(i[0])+' '+str(i[1])+' '+str(i[2])+' '+1.0'
234             PH4.write(cur_str)
235             PH4.write("\n")
236         f = '#volume '+str(EXCL_nr)+' size i expr tt color ix ebits ix gbits ix'
237         PH4.write(f)
238         PH4.write("\n")
239         for i in range(EXCL_nr):
240             EXCL_str = '1 $ 606060 1 300'
241             PH4.write(EXCL_str)
242             PH4.write("\n")
243         ### END TAG
244         end = '#endpharmacophore'
245         PH4.write(end)
246         PH4.close()
247         print "CCG ph4 successfully written!", time.ctime()
248         print "\n"
249         ### path variables
250         pharao = '/opt/pharao-3.0.3/pharao'
251         ### Main programm
252         start = time.ctime()
253         start_dir = os.getcwd()
254         all_files = os.listdir(start_dir)
255         all_folders = [name for name in all_files if not '.py' in name and not '.txt' in name]
256         all_folders.sort()
257         for target in all_folders:
258             os.chdir(target)
259             sub_dir = os.getcwd()
260             sub_files = os.listdir(sub_dir)
261             ### DEFINE PHARMACOPHORE FILES
262             ### visible mol
263             mol = openbabel.OBMol()
264             SB_ph4_sdf = target+'_struct_based_phar.sdf'
265             ### INTERACTION BASED PH4
266             SB_ph4_name = '2_'+target+'_struct_based.phar'
267             SB_ph4_header = '2_'+target+'_struct_based'
268             SB_ph4 = file(SB_ph4_name, 'w')
269             SB_ph4.write(SB_ph4_header)
270             SB_ph4.write("\n")
271             ### create CCG ph4 files here!
272             ### ENERGETIC PH4
273             ### 04-feb-2012 (fk)
274             SB_CCG_ph4_e_name = '3_'+str(target)+'_struct_based_CCG_E.ph4'
275             SB_CCG_ph4_ENERGETIC = file(SB_CCG_ph4_e_name, 'w')
276             ### INTERACTION BASED
277             SB_CCG_ph4_i_name = '4_'+str(target)+'_struct_based_CCG.ph4'
278             SB_CCG_ph4_INTERACT = file(SB_CCG_ph4_i_name, 'w')
279             ### Ligbased + EXCL
280             SB_ph4_name_III = '5_'+target+'_ligbased_EX.phar'
281             SB_ph4_header_III = '5_'+target+'_ligbased_EX'
282             SB_ph4_III = file(SB_ph4_name_III, 'w')
283             SB_ph4_III.write(SB_ph4_header_III)
284             SB_ph4_III.write("\n")
285             ### ph4 unit (ligand-based)
286             ref_ligand = [name for name in sub_files if 'refined_LIGAND_3D.mol' in name][0]
287             ph4_ligbased = create_ph4(ref_ligand, target)
288             AROM = ph4_ligbased[0]
289             HDON = ph4_ligbased[1]
290             HACC = ph4_ligbased[2]
291             LIPO = ph4_ligbased[3]
292             POSC = ph4_ligbased[4]
293             NEGCG = ph4_ligbased[5]
294             print "ph4 info successfully read out!"
295             print "\n"
296             lig_coords = get_lig_coords_mol(ref_ligand)
297             ### 16-may-2011

```

```

298 lig_dicts = dict_the_lig(ref_ligand)
299 coord_2_idx = lig_dicts[0]
300 idx_2_coord = lig_dicts[1]
301 ### protein unit
302 ref_protein = [name for name in sub_files if 'refined_PROTEIN_3D.pdb' in name][0]
303 protein_data = processing_protein(ref_protein)
304 structure = protein_data[0]
305 resdict = protein_data[1]
306 ### dict 4 CCG interacting ph4_data
307 ### contact
308 CCG_interact = []
309 ### energetic
310 CCG_interact_E = []
311 ### interaction unit
312 ### LIPO
313 if len(LIPO)>0:
314     LIPO_interact = sift_HYD(LIPO, structure, rules=SDRY ,RANGE=[2.3,6.0], Min_contact_res=3)
315
316     if len(LIPO_interact)>0:
317         written_LIPO = write_SB_ph4(SB_ph4, LIPO_interact, prefix='LIPO', write=True)
318         for i in LIPO_interact:
319             i.append('LIPO')
320             CCG_interact.append(i)
321         ### --> E LIPO
322         lipo_hotkont = [name for name in sub_files if 'C3_HOT_kont.txt' in name][0]
323         LIPO_E = read_HOTKONT(lipo_hotkont, LIPO, feat_type='LIPO')
324         if len(LIPO_E)> 0:
325             for i in LIPO_E:
326                 CCG_interact_E.append(i)
327 ### AROM
328 if len(AROM)>0:
329     AROM_interact = sift_HYD(AROM, structure, rules=SDRY ,RANGE=[2.3,6.0], Min_contact_res=3)
330     ### separate function for pi-pi; 11-sept-2011
331     PI_PI_interact = sift_PI_PI(AROM, structure, rules=AROMATIC ,RANGE=[3.0, 6.0])
332     for i in PI_PI_interact:
333         if i not in AROM_interact:
334             AROM_interact.append(i)
335         else:
336             print "AROM interaction already recognized!"
337     if len(AROM_interact)>0:
338         written_AROM = write_SB_ph4(SB_ph4, AROM_interact, prefix='AROM', write=True)
339         for i in AROM_interact:
340             i.append('AROM')
341             CCG_interact.append(i)
342         ### --> E AROMATIC
343         arom_hotkont = [name for name in sub_files if 'C3_HOT_kont.txt' in name][0]
344         AROM_E = read_HOTKONT(arom_hotkont, AROM, feat_type='AROM')
345         if len(AROM_E)> 0:
346             for i in AROM_E:
347                 CCG_interact_E.append(i)
348 ### NEGC
349 if len(NEGC)>0:
350     NEGC_interact = sift_NEGC(NEGC, structure, rules=POSC_rules, RANGE=[2.3, 4.5])
351     if len(NEGC_interact)>0:
352         written_NEGC = write_SB_ph4(SB_ph4, NEGC_interact, prefix='NEGC', write=True)
353         for i in NEGC_interact:
354             i.append('NEGC')
355             CCG_interact.append(i)
356         ### --> E NEGC
357         negc_hotkont = [name for name in sub_files if '_O-_HOT_kont.txt' in name][0]
358         print negc_hotkont
359         NEGC_E = read_HOTKONT(negc_hotkont, NEGC, feat_type='NEGC')
360         if len(NEGC_E)> 0:
361             for i in NEGC_E:
362                 CCG_interact_E.append(i)
363 ### DONOR
364 if len(HDON) > 0:
365     HDON_interact = sift_HDON(HDON, structure, coord_2_idx, idx_2_coord, ref_ligand, rules=
366         dict_acceptor,
367         rules_II = NEGC_rules, RANGE=[2.3,3.5])
368     if len(HDON_interact)>0:
369         written_HDON = write_SB_ph4(SB_ph4, HDON_interact, prefix='HDON', write=True)
370         for i in HDON_interact:
371             i.append('HDON')
372             CCG_interact.append(i)
373         ### --> E NEGC
374         hdon_hotkont = [name for name in sub_files if '_N1_HOT_kont.txt' in name][0]

```

```

374 HDON_E = read_HOTKONT(hdon_hotkont,HDON, feat_type='HDON')
375 if len(HDON_E)> 0:
376     for i in HDON_E:
377         CCG_interact_E.append(i)
378 ### POSC
379 if len(POSC)>0:
380     POSC_interact = sift_POSC(POSC,structure, rules= NEGC_rules, RANGE=[2.3, 4.5])
381     CAT_PI_interact = sift_CAT_PI(POSC,structure, rules=AROMATIC ,RANGE=[2.3, 4.5])
382     for i in CAT_PI_interact:
383         if i not in POSC_interact:
384             POSC_interact.append(i)
385     else:
386         print "POSC interaction already recognized!"
387 if len(POSC_interact)>0:
388     written_POSC = write_SB_ph4(SB_ph4, POSC_interact, prefix='POSC', write=True)
389     for i in POSC_interact:
390         i.append('POSC')
391     CCG_interact.append(i)
392 ### --> E POSC
393 posc_hotkont = [name for name in sub_files if '_Nl+_HOT_kont.txt' in name][0]
394 POSC_E = read_HOTKONT(posc_hotkont,POSC, feat_type='POSC')
395 if len(POSC_E)> 0:
396     for i in POSC_E:
397         CCG_interact_E.append(i)
398 ### ACCEPTOR
399 if len(HACC)> 0:
400     HACC_interact = sift_HACC(HACC, structure, ref_ligand, rules=dict_donor, RANGE = [2.3,3.5])
401     if len(HACC_interact)>0:
402         written_HACC =write_SB_ph4(SB_ph4, HACC_interact, prefix='HACC', write=True)
403         for i in HACC_interact:
404             i.append('HACC')
405         CCG_interact.append(i)
406 ### --> E ACCEPTOR
407 hacc_hotkont = [name for name in sub_files if '_O+_HOT_kont.txt' in name][0]
408 HACC_E = read_HOTKONT(hacc_hotkont,HACC, feat_type='HACC')
409 if len(HACC_E)> 0:
410     for i in HACC_E:
411         CCG_interact_E.append(i)
412 ### set cavity flanking res as EXCL
413 EXCL = create_EXCL(structure ,lig_coords, alpha= 1.0)
414 written_excl = write_SB_ph4(SB_ph4, EXCL, prefix='EXCL', write=False)
415 ### CCG ph4 files ...
416 ### --> create CCG ph4 PPCD
417 ### interaction based
418 SB_CCG_Interact = write_CCG_ph4(CCG_interact, EXCL, PH4=SB_CCG_ph4_INTERACT, flag='I')
419 ### energy based
420 CCG_interact_E.sort()
421 CCG_interact_E.reverse()
422 ### take the best !! (rules according to Salam et. al. )
423 if len(CCG_interact_E) >=4:
424     best_CCG_interact_E = CCG_interact_E[0:5]
425 if len(CCG_interact_E) >=5:
426     best_CCG_interact_E = CCG_interact_E[0:6]
427 else:
428     best_CCG_interact_E = CCG_interact_E
429 SB_CCG_Energetic = write_CCG_ph4(best_CCG_interact_E, EXCL, PH4=SB_CCG_ph4_ENERGETIC, flag='
E')
430 ### finish pharao (.phar) files ...
431 pybelmol = pybel.Molecule(mol)
432 pybelmol.write("sdf", SB_ph4_sdf, overwrite = True)
433 SB_ph4.write("$$$$")
434 SB_ph4.close()
435 ### write Ligbased + EXCL
436 sub_files = os.listdir(sub_dir)
437 ligbased_phar = [name for name in sub_files if '_ligbased.phar' in name][0]
438 lig_ex_phar = write_ligbased_EX(SB_ph4_III,ligbased_phar,EXCL)
439 SB_ph4_III.write("$$$$")
440 SB_ph4_III.close()
441 os.chdir(start_dir)
442 print "done.", time.ctime()
443 print "\n"

```

Kapitel 27

Parameter und Konfigurationsdateien

27.1 omega_HTS.parm (OMEGA)

```
#InterfACE settings
#-pvmconf
#File Options :
-commentEnergy true
#-in
#-includeInput false
#-log
#-out out.sdf
#-param
-prefix omega
-rotorOffsetCompress false
-sdEnergy true
#-status
-verbose false
-warts true
#3D Construction Parameters :
-buildff mmff94s_Trunc
-canonOrder true
-deleteFixHydrogens true
#-dielectric
#-exponent
#-fixfile
-fixrms 0.150000
-fraglib /home/flokoell/PARAM/fraglib.oeb.gz
-fromCT true
-maxmatch 100
-umatch true
#Structure Enumeration :
-enumNitrogen true
-enumRing true
#Torsion Driving Parameters :
#-erange
-ewindow 25.000000
#-maxConfRange 50,200,400
-maxconfgen 30000
-maxconfs 100
-maxpoolsize 10000
-maxrot -1
-maxtime 10.0
-rangeIncrement 5
-rms 1
#-rmsrange 0.8,1,1.2
-searchff mmff94s_Trunc
-tordrive true
#-torlib
#PVM Parameters :
#-pvmdebug true
#-pvmpass 50
#-pvmconf /home/flokoell/PARAM/hosts_pvm
```


27.2 omega_HQS.parm (OMEGA)

```

#InterfACE settings
#-pvmconf
#File Options :
-commentEnergy true
#-in
#-includeInput false
#-log
#-out out.sdf
#-param
-prefix omega2
-rotorOffsetCompress false
-sdEnergy true
#-status
-verbose true
-warts true
#3D Construction Parameters :
-buildff mmff94s_Trunc
-canonOrder true
-deleteFixHydrogens true
#-dielectric
#-exponent
#-fixfile
-fixrms 0.150000
-fraglib /home/flokoell/PARAM/fraglib.oeb.gz
-fromCT true
-maxmatch 400
-umatch true
#Structure Enumeration :
-enumNitrogen true
-enumRing true
#Torsion Driving Parameters :
#-erange
-ewindow 20.000000
#-maxConfRange (5,10,150,200)
-maxconfgen 10000
-maxconfs 500
-maxpoolsize 1000
-maxrot -1
-maxtime 100.000000
-rangeIncrement 5
-rms 0.600000
#-rmsrange
-searchff mmff94s_Trunc
-tordrive true
#-torlib
#PVM Parameters :
#-pvmdebug true
#-pvmpass 20

#-pvmconf /home/flokoell/PARAM/hosts_pvm

```

27.3 tautomers.param (QUACPAC)

```
#InterfACE settings
#tautomer :
#-all false
-can true
#-ch3 false
#-count false
-in SRC__decoys_pka.sdf
#-kekule false
#-level 0
#-max 1000
-out SRC__decoys_pka_tautomers.sdf
#-param (Not set, no default)
#-paramfile (Not set, no default)
#-prefix tautomers
-reasonable true

#-uniq false
```

27.4 pkatyper.param (QUACPAC)

```
#InterfACE settings
#pkatyper :
#-count false
-in SRC__decoys.sdf
#-max 100
-out SRC__decoys_pka.sdf
#-param (Not set, no default)
#-paramfile (Not set, no default)

#-prefix pkatyper
```

27.5 GRID Parameter

Im folgenden sind zwei beispielhafte GRID-Parameter-Dateien *grin.in* und *grid.in* des *Angiotensin-Konversionsenzyms* (Abk. ACE; PDB: 1O86) unter Verwendung der Wasser-Sonde (H₂O) gezeigt. Das GRID-Unterprogramm *grin* liest das von dem Programm 'prepare_grid_cavka.py' (vgl. Abschn. 25.2.2) vorbereitete Protein im .pdb Format ein, definiert Wasserstoffpositionen und bereitet die GRID-Berechnungen vor.

Im Rahmen einer GRID-Berechnung werden die folgenden Dateien für jede Struktur pro Sonde generiert:

- *file.list* (enthält nur den Dateinamen)
- *.lout* (wird von dem Programm *grin* erzeugt und listet nicht eindeutig bezeichnete Aminosäure-Reste in der .pdb Datei auf)
- *.kout* (enthält Informationen der durch *grin* prozessierten .pdb Datei und dient als Input für die GRID-Berechnungen)
- *.lont* (enthält die bei den GRID-Berechnungen anfallende Kommandozeilen-Ausgabe)
- *.kont* (enthält die berechneten Interaktionswerte der verwendeten Sonde sowie die zugehörigen Koordinaten in einem binären-, bzw. lesbaren Ascii-Format. Die .kont Dateien bilden die eigentliche Grundlage für die in CavKA berechneten MIFs).

27.5.1 grin.in

```
LOUT 6
LOUT 1O86_OH2.lout
KOUT 20
KOUT 1O86_OH2.kout
INKO 11
INKO 1O86_protein_NO_H_clean.pdb
INAT 10
INAT /opt/grid/grub.dat
ALHY 0.040
NEHY 1
IHVA 0
LEVL 3
MOVE 1
QQHY 0.000
VDHY 0.600

IEND
```

27.5.2 grid.in

```
LONT 6
LONT 1086_OH2.lont
KONT 20
KONT 1086_OH2.kont
INPT 10
INPT file.list
CLER 5.000
DEEP 5.000
DPRO 4.000
DWAT 80.000
EACH 5.000
EMAX 5.000
FARH 5.000
FARR 8.000
ALMD 0.000
KWIK 1
LEAU 0
LENG 12
LEVL 1
LIST 1
MOVE 0
NETA 0
NPLA 2
NUMB 1
VALU 0.000
TOPX 49.000
TOPY 44.000
TOPZ 59.000
BOTX 32.000
BOTY 22.000
BOTZ 36.000
OH2
IEND
TEST
```

```
0 1
```

Kapitel 28

Daten CavKA: Harte-Sphären-Modelle

28.1 Vergleich von CavKA mit Ligandscout

Tabellarische Zusammenfassung der in Kapitel 18 diskutierten Ergebnisse. Die aufgeführten Pharmakophor-Modelle sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.

Tabelle 28.1: Screening-Ergebnisse (*Sensitivität* und *Spezifität*) des Vergleichs von CavKA und Ligandscout unter Angabe der manuell-optimierten Pharmakophor-Modelle auf dem Fieldscreen-Datensatz¹⁸⁴ (verwendete PDB-Struktur in Klammern; Datensatz-Vorbehandlung: struct_HTS).

CavKA Modell	Sensi- tivität	Spezi- fität	Ziel- Struktur	Sensi- tivität	Spezi- fität	Ligandscout Modell
5_ACE_1o86_struct_ based_CCG_i.ph4	0.88	0.92	ACE (1o86)	0.67	0.94	8_1o86_LIGAND- SCOUT_PCHD.ph4
8_ache_1eve_struct_based_ based_CCG_i_WATER.ph4	0.42	0.78	AChE (1eve)	0.18	0.90	12_ache_LIGAND- SCOUT_PCHD.ph4
6_CDK2_1ckp_struct_ based_CCG.ph4	0.33	0.79	CDK2 (1ckp)	0.80	0.17	11_CDK2_LIGAND- SCOUT_PCHD.ph4
5_COX2_1cx2_struct_ based_CCG_i.ph4	0.44	0.92	COX2 (1cx2)	0.48	0.95	12_COX2_1cx2_LIGAND- SCOUT_PCHD.ph4
4_EGFR_1m17_struct_ based_CCG_i.ph4	0.52	0.76	EGFR (1m17)	0.78	0.30	10_EGFR_1m17_LIGAND- SCOUT.ph4
4_FXa_1f0r_struct_ based_CCG.ph4	0.05	0.98	FXa (1f0r)	0.26	0.75	10_FXa_1f0r_LIGAND- SCOUT_PCHD.ph4

Tabelle 28.2: (Fortsetzung von Tab. 28.1). Screening-Ergebnisse (*Sensitivität* und *Spezifität*) des Vergleichs zwischen CavKA und Ligandscout unter Angabe der manuell optimierten Pharmakophor-Modelle auf dem Fieldscreen-Datensatz¹⁸⁴ (verwendete PDB-Struktur in Klammern; Datensatz-Vorbehandlung: struct_HTS).

CavKA Modell	Sensitivität	Spezifität	Ziel-Struktur	Sensitivität	Spezifität	Ligandscout Modell
6_HIV-RT_1rt1_struct_based_CCG_i.ph4	0.12	0.88	HIV-RT (1rt1)	-	-	-
4_INHA_1p44_struct_based_CCG.ph4	0.45	0.54	INHA (1p44)	0.23	0.97	9_INHA_1p44_LIGANDSCOUT.ph4
6_P38_1kv2_struct_based_CCG_i.ph4	0.90	0.31	P38 (1kv2)	0.57	0.55	12_P38_1kv2_struct_based_LIGANDSCOUT_PCHD.ph4
8_PDE5_1xp0_struct_based_CCG.ph4	0.40	0.86	PDE5 (1xp0)	0.08	0.95	9_PDE5_1xp0_struct_based_LIGANDSCOUT.ph4
6_PDGFRB_1t46_struct_based_CCG.ph4	0.36	0.98	PDGFRB (1t46)	-	-	-
-	-	-	SRC (2SRC)	0.02	0.84	6_SRC_2SRC_struct_based_LIGANDSCOUT.ph4
4_VEGFR2_1fgi_struct_based_CCG.ph4	0.06	0.96	VEGFR2 (1fgi)	0.06	0.96	11_VEGFR2_1fgi_struct_based_LIGANDSCOUT.ph4

28.2 Einfluss der Konformeren-Auflösung

Tabellarische Zusammenfassung der in Abschnitt 18.3 diskutierten Ergebnisse. Die aufgeführten Pharmakophor-Modelle sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.

Tabelle 28.3: Einfluss der Konformeren-Auflösung (Datensatz-Vorbehandlung: struct_HTS und struct_HQS) auf die Screening-Ergebnisse (*Sensitivität* und *Spezifität*) für die manuell optimierten Modelle (CavKA, bzw. Ligandscout) auf dem Fieldscreen-Datensatz¹⁸⁴ (verwendete PDB-Struktur in Klammern).

Ziel-Struktur	Bestes Pharmakophor-Modell	Sens. (HTS)	Spez. (HTS)	Sens. (HQS)	Spez. (HQS)
ACE (1o86)	4_ACE_1o86_struct_ based_CCG_i.ph4	0.88	0.91	0.88	0.90
AChE (1eve)	8_AChE_1eve_struct_ based_CCG_i_WATER.ph4	0.42	0.78	0.42	0.74
CDK2 (1ckp)	6_CDK2_1ckp_struct_ based_CCG.ph4	0.33	0.79	0.33	0.76
COX2 (1cx2)	12_COX2_1cx2_ LIGANDSCOUT_PCHD.ph4	0.48	0.95	0.51	0.94
EGFR (1m17)	4_EGFR_1m17_struct_ based_CCG_i.ph4	0.52	0.76	0.53	0.76
FXa (1f0r)	4_FXa_1f0r_struct_ based_CCG.ph4	0.05	0.98	0.05	0.97
HIV-RT (1rt1)	6_HIV-RT_1rt1_struct_ based_CCG_i.ph4	0.12	0.88	0.20	0.84
INHA (1p44)	9_INHA_1p44_ LIGANDSCOUT.ph4	0.23	0.97	0.23	0.96
P38 (1kv2)	7_P38_1kv2_ GLIDE_XP.ph4	0.28	0.81	0.45	0.78
PDE5 (1xp0)	8_PDE5_1xp0_struct_ based_CCG.ph4	0.40	0.86	0.44	0.84
PDGFRB (1t46)	6_PDGFRB_1t46_struct_ based_CCG.ph4	0.36	0.98	0.41	0.98
SRC (2SRC)	6_SRC_2SRC_struct_ based_LIGANDSCOUT.ph4	0.02	0.84	0.05	0.79
VEGFR2 (1fgi)	5_VEGFR2_1fgi_struct_ based_CCG.ph4	0.06	0.96	0.06	0.95

28.3 Leistungsfähigkeit von E-Pharmakophoren im Vergleich

Tabellarische Zusammenfassung der in Kapitel 19 diskutierten Ergebnisse. Die aufgeführten Pharmakophor-Modelle sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert. Die sortierten Hitlisten sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.

Tabelle 28.5: *Sensitivität* (Abk. Sens.) und *Spezifität* (Abk. Spez.) von E-Pharmakophor-Modellen¹⁶⁴ (Abk. E-Phar) und MIF-basierten-E-Pharmakophor-Modellen (Abk. E-MIF) im Vergleich mit dem besten manuell optimierten Pharmakophor-Modell (CavKA, bzw. Ligandscout; Abk. Best) auf dem Fieldscreen¹⁸⁴-Datensatz (verwendete PDB-Struktur in Klammern; Datensatz-Vorbehandlung: struct_HTS).

Ziel-Struktur	Sens. (E-Phar)	Spez. (E-Phar)	Sens. (E-MIF)	Spez. (E-MIF)	Sens. (Best)	Spez. (Best)
AChE (1eve)	0.48	0.39	0.46	0.63	0.42	0.78
COX2 (1cx2)	1.00	0.02	0.17	0.96	0.48	0.95
EGFR (1m17)	0.60	0.65	0.17	0.89	0.52	0.76
HIV-RT (1rt1)	0.03	0.90	0.52	0.34	0.12	0.88
P38 (1kv2)	0.28	0.81	-	-	0.57	0.55

Tabelle 28.6: Verwendete Pharmakophor-Modelle für den Vergleich der Leistungsfähigkeit von E-Pharmakophor-Modellen¹⁶⁴ (Abk. E-Phar), MIF-basierten-E-Pharmakophor-Modellen und dem besten gefundenen Modell (methodenunabhängig; Abk. BEST; vgl. Tab. 28.5).

Ziel-Struktur	Modell (E-Phar)	Modell (E-MIF)	Modell (Best)
AChE (1eve)	13_AChE_ GLIDE_XP.ph4	3_AChE_1eve_struct_ based_CCG_E.ph4	8_AChE_1eve_struct_ based_CCG_i_WATER.ph4
COX2 (1cx2)	8_COX2_1cx2_struct_ based_GLIDE_xp.ph4	3_COX2_1cx2_struct_ based_CCG_E.ph4	12_COX2_1cx2_ LIGANDSCOUT_PCHD.ph4
EGFR (1m17)	7_EGFR_1m17_ GLIDE_XP.ph4	3_EGFR_1m17_struct_ based_CCG_E.ph4	4_EGFR_1m17_struct_ based_CCG_i.ph4
HIV-RT (1rt1)	10_HIV-RT_1rt1_ GLIDE_XP.ph4	2_HIV-RT_1rt1_struct_ based_CCG_E.ph4	6_HIV-RT_1rt1_struct_ based_CCG_i.ph4
P38 (1kv2)	7_P38_1kv2_ GLIDE_XP.ph4	-	12_P38_1kv2_struct_ LIGANDSCOUT_PCHD.ph4

Kapitel 29

PLIDriPaS

Tabellarische Zusammenfassung der in Kapitel 20 diskutierten Ergebnisse. Die aufgeführten Pharmakophor-Modelle sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.

Tabelle 29.1: *Sensitivität* (Abk. Sens.) und *Spezifität* (Abk. Spez.) für die auf Basis von zehn sc-PDB²³¹-Ligand-Rezeptor-Komplexen erstellten Pharmakophor-Modelle. *Modell* nennt die manuell optimierten Pharmakophor-Modelle mit möglichst hoher Anreicherung bei möglichst hoher Spezifität. Fett gedruckte Modelle sind für die Analysen in Kapitel 15 verwendet worden (Datensatz-Vorbehandlung: struct_HTS).

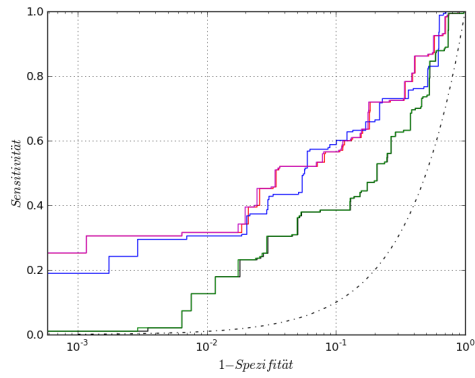
Cluster	PDB	Ligand	Sens.	Spez.	Modell
(I)	1uho	Vardenafil	0.59	0.78	7_1uho_cavka.ph4
(I)	1xp0	Vardenafil	0.18	0.95	9_1xp0.ph4
(I)	1tbf	Sildenafil	0.23	0.97	3_1tbf_cavka.ph4
(I)	2h42	Sildenafil	0.05	0.99	11_2h42_cavka.ph4
(I)	3jwq	Sildenafil	0.32	0.96	16_3jwq_cavka.ph4
(I)	3bjc	5-ethoxy-4-(1-methyl-7-oxo-3-propyl-6,7-dihydro-1H-pyrazolo[4,3-d]pyrimidin-5-yl)thiophene-2-sulfonamide	0.05	0.99	14_3bjc_cavka.ph4
(II)	1t9s	Guanosin-5'-Monophosphat	0.00	1.00	1t9s_cavka.ph4
(III)	1xoz	Tadalafil	0.32	0.98	8_1xoz.ph4
(IV)	2h44	Icarisid II	0.05	0.98	13_2h44_cavka.ph4
(V)	3jwr	3-Isobutyl-1-Methoxyxanthin	0.32	0.87	20_3jwr.ph4

Kapitel 30

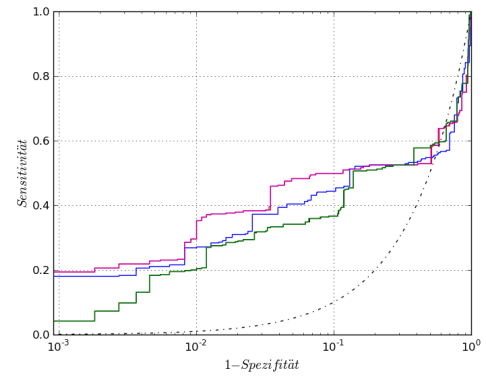
Daten: CavKA^{*HYBRID*}

30.1 CavKA^{*HYBRID*}: ROC-Kurven

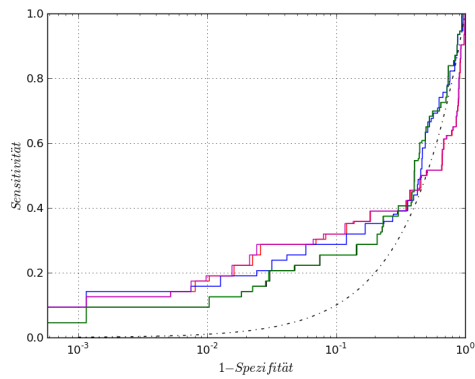
ROC-Kurven zu den in Kapitel 21 diskutierten Ergebnissen. Die verwendeten Pharmakophor-Modelle sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.



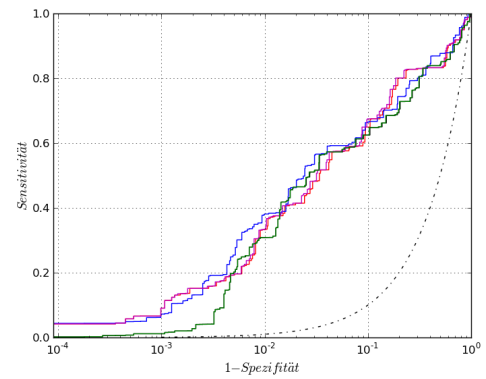
(a) ACE



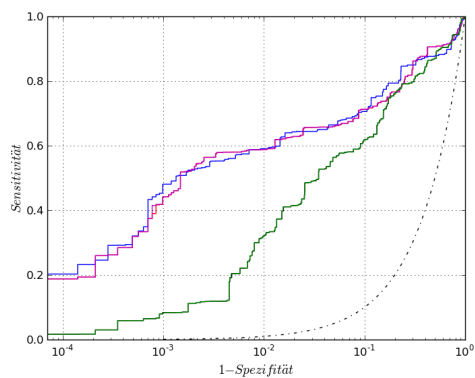
(b) AChE



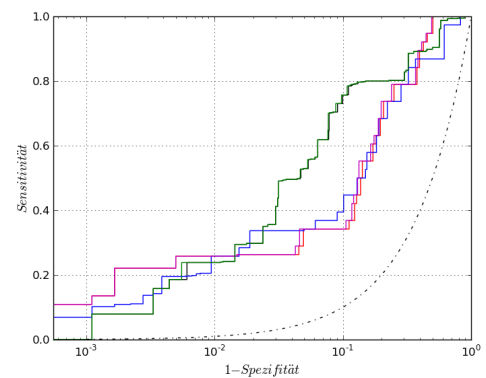
(c) CDK2



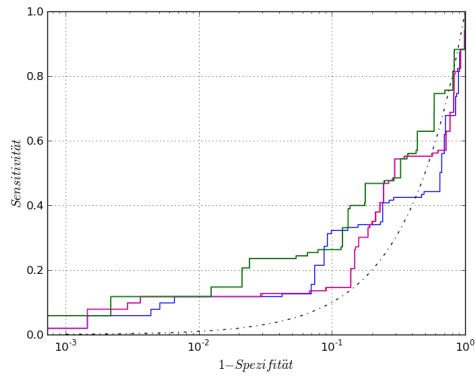
(d) COX2



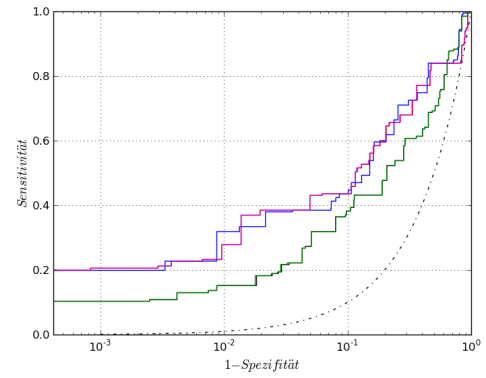
(e) EGFR



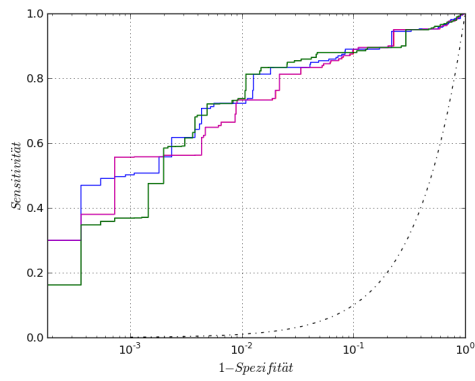
(f) FXa



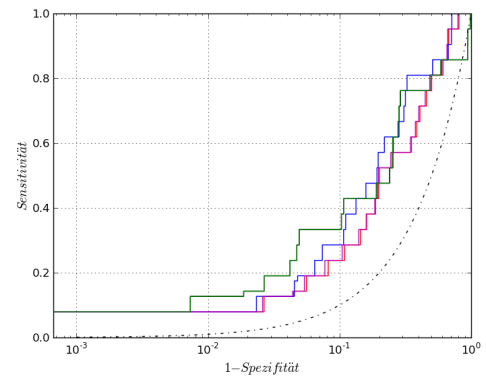
(g) HIV-RT



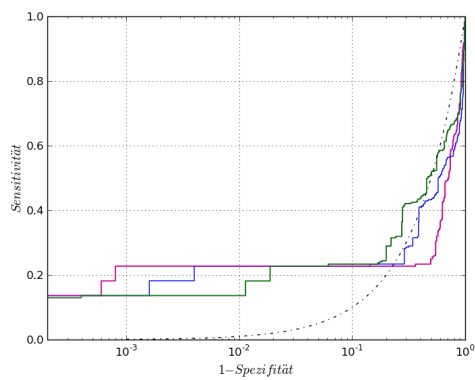
(h) INHA



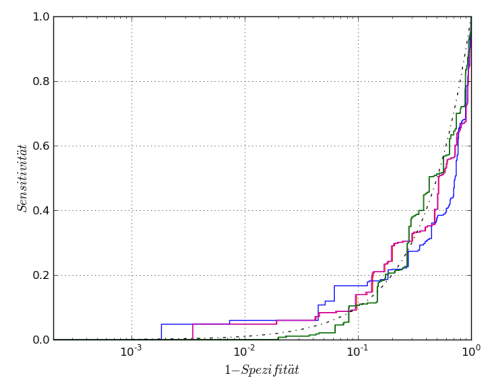
(i) P38



(j) PDE5



(k) PDGFRB



(l) SRC

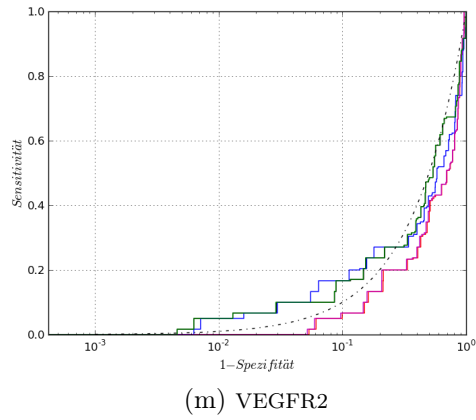


Abbildung 30.1: ROC-Kurven für den Vergleich von Pharao und CavKA^{HYBRID}. Rot: ligandzentrisch; schwarz: ligandzentrisch + Gaussian; blau: strukturbasiert + Gaussian; magenta: ligandzentrisch^{HYBRID}; grün: CavKA^{HYBRID}; schwarz gepunktet: Zufall. Datensatz-Vorbehandlung: struct_HTS.

30.2 CavKA^{HYBRID}: ROC-Anreicherungs-faktoren

Grafiken mit $ROC_{EF5\%}$ -Werten und $ROC_{EF510\%}$ -Werten zu den in Kapitel 21 diskutierten Ergebnissen.

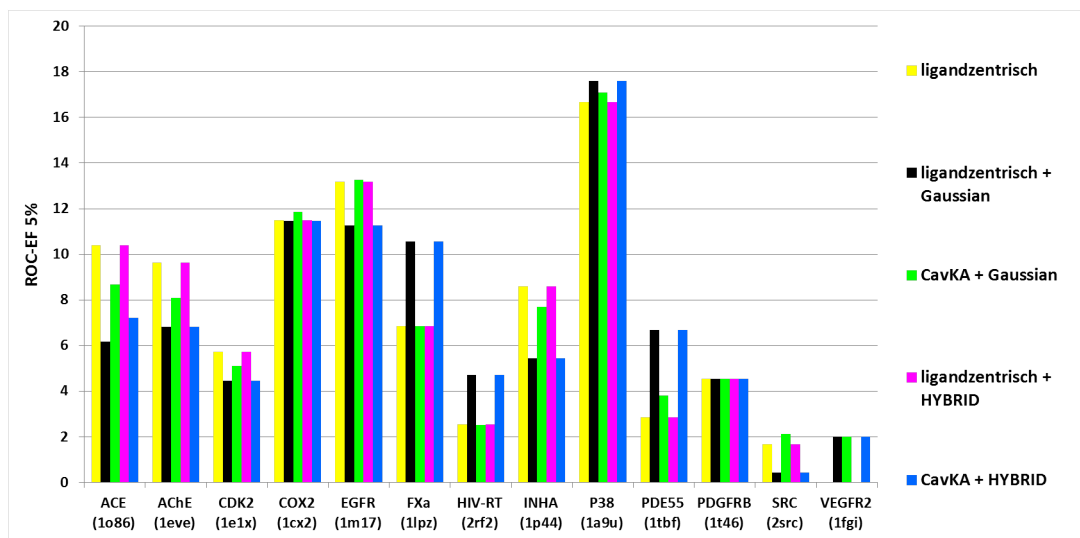


Abbildung 30.2: $ROC_{EF5\%}$ -Werte für die 13 Fieldscreen Zielstrukturen. Die verwendeten Pharmakophor-Modelle wurden nach unterschiedlichen Methoden abgeleitet (s. Abschn.: 13.3; Datensatz-Vorbehandlung: struct_HTS).

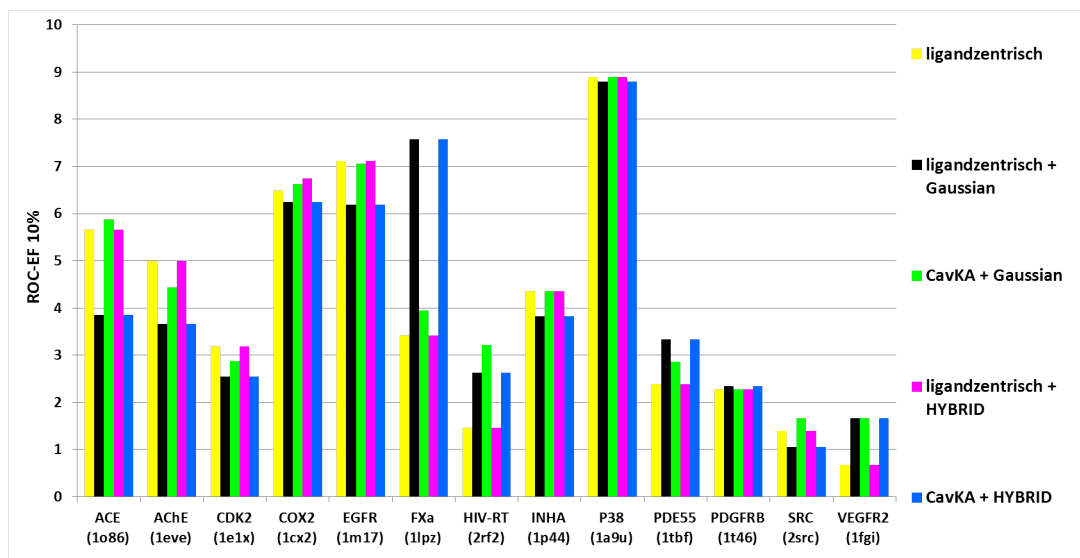


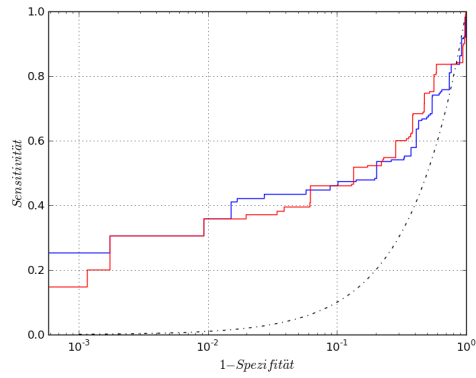
Abbildung 30.3: $ROC_{EF10\%}$ -Werte für die 13 Fieldscreen Zielstrukturen. Die verwendeten Pharmakophor-Modelle wurden nach unterschiedlichen Methoden abgeleitet (s. Abschn.: 13.3; Datensatz-Vorbehandlung: struct_HTS).

Kapitel 31

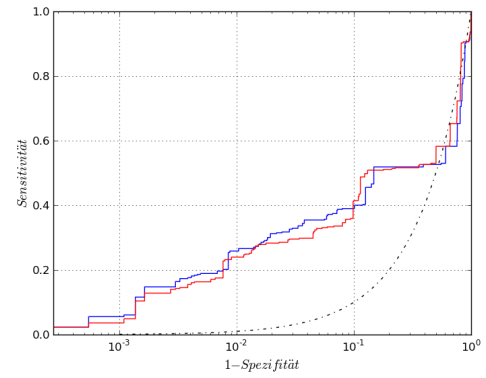
Daten: Ligandzentrische Ansätze

31.1 Einfluss unterschiedlicher ROCS-Superpositionierungen auf die Screening- Ergebnisse mit EON

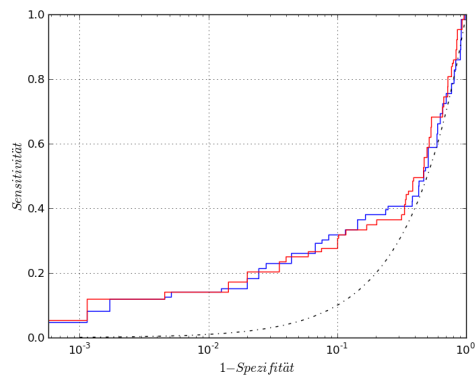
ROC-Kurven für Screening mit EON unter Verwendung unterschiedlicher ROCS-Superpositionierungen (vgl. Abschn. 22.3.2). Die sortierten Hitlisten sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.



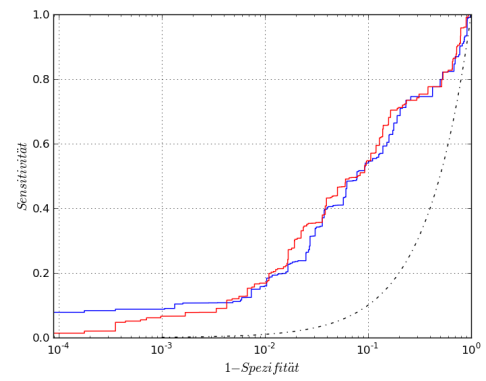
(a) ACE



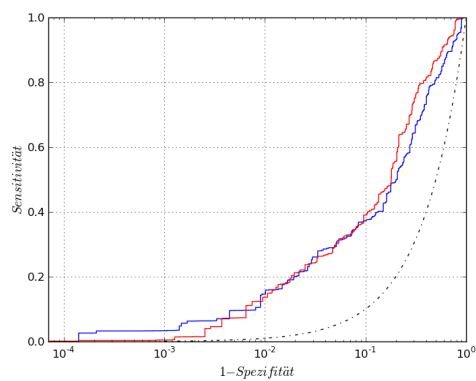
(b) AChE



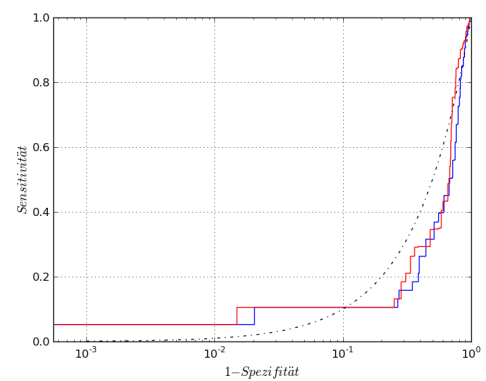
(c) CDK2



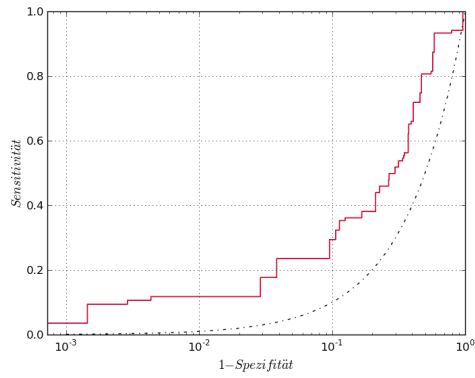
(d) COX2



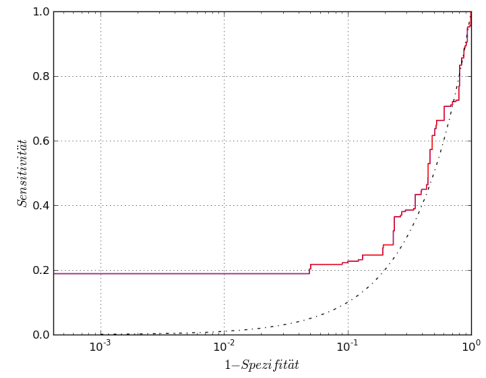
(e) EGFR



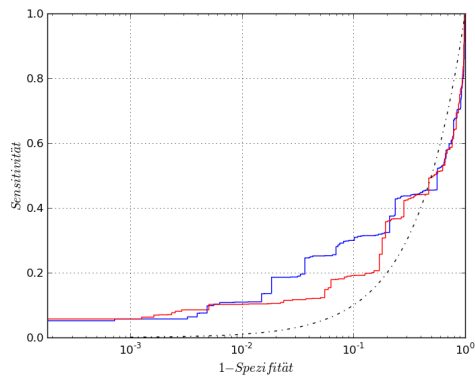
(f) FXa



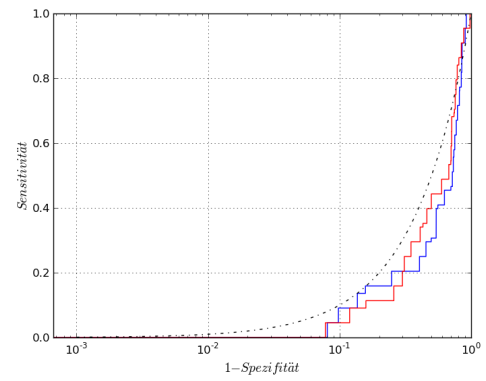
(g) HIV-RT



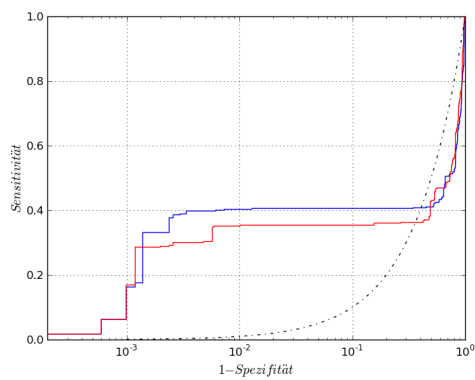
(h) INHA



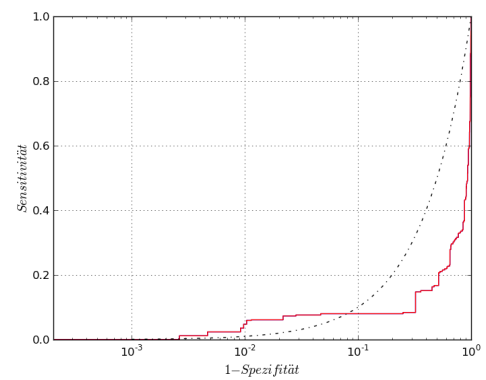
(i) P38



(j) PDE5



(k) PDGFRB



(l) SRC

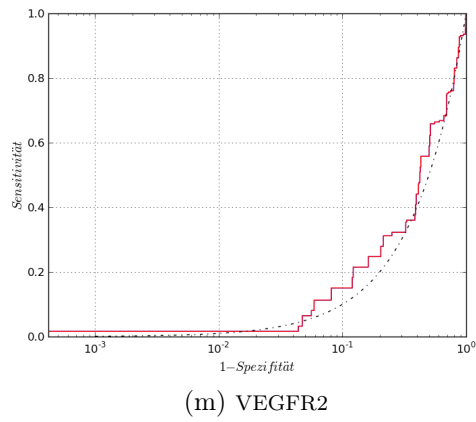


Abbildung 31.1: Einfluss unterschiedlicher ROCS Superpositionierungen auf die Screening-Ergebnisse mit EON. Rot: $ROCS_{combo}$ -Superpos.; blau: $ROCS - EON$ -Superpos.; schwarz gepunktet: Zufall. Als Scoring-Funktion für die jeweiligen Superpositionierungen wurde EON_{combo} verwendet (Datensatz-Vorbehandlung: *ligcent_HTS*).

31.2 Alternative P38-Referenz-Liganden

ROC-Kurven, die unter Verwendung unterschiedlicher P38-Referenz-Liganden erhalten worden sind (vgl. Diskussion in Abschnitt 18.2.3). Die sortierten Hitlisten sind in digitaler Form am Institut für Medizinische und Pharmazeutische Chemie archiviert.

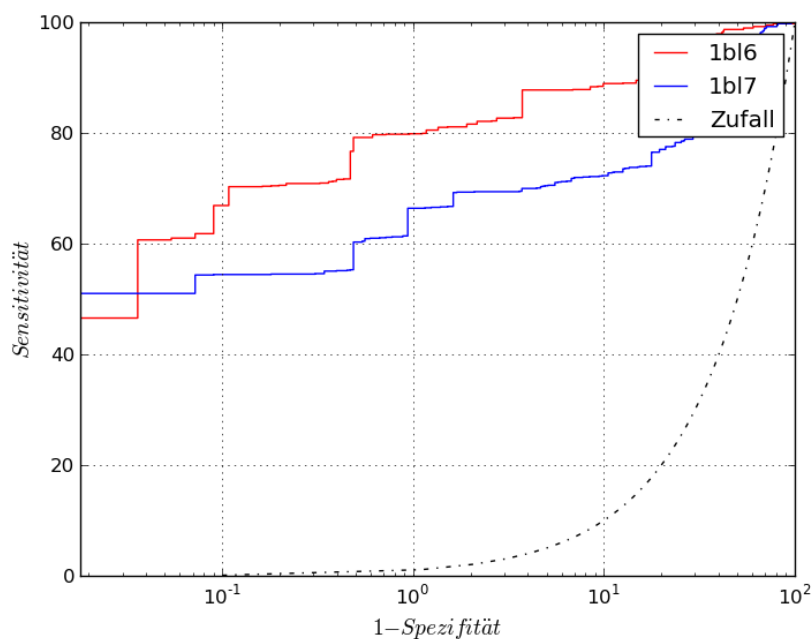


Abbildung 31.2: Alternative P38-Referenz-Liganden. Wird der co-kristallisierte Ligand aus PDB 1bl6 auf dem Fieldscreen-Datensatz im Rahmen eines Screenings verwendet, werden höhere Anreicherungen erzielt als mit der von KINNINGS und JACKSON vorgeschlagenen Referenz (der co-kristallisierte Ligand aus PDB 1bl7) ²⁶¹. *Datensatz-Vorbehandlung: ligcent_HTS*.

Literaturverzeichnis

- [1] Klebe, G. *Wirkstoffdesign*, 2nd ed.; Spektrum Verlag, 2009.
- [2] Klebe, G. Lead identification in post-genomics: Computers as a complementary alternative. *Drug Discovery Today* **2004**, *1*, 225–230.
- [3] Schnecke, V.; Bostroem, J. Computational chemistry-driven decision making in lead generation. *Drug Discovery Today* **2006**, *11*, 43–50.
- [4] Aheme, G. W.; Mc Donald, E.; Workman, P. Finding the needle in the haystack: Why high-throughput screening is good for your health. *Breast Cancer Res.* **2002**, *4*, 148–154.
- [5] Mayr, L. M.; Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **2009**, *9*, 580–588.
- [6] Sundberg, S. A. High-throughput and ultra-high-throughput screening: Solution- and cell-based approaches. *Curr. Opin. Biotechnol.* **2000**, *11*, 47–53.
- [7] Malo, N.; Hanley, J. A.; Cerquozzi, S.; Pelletier, J.; Nadon, R. Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.* **2006**, *24*, 167–175.
- [8] Inglese, J.; Auld, D. S.; Jadhav, A.; Johnson, R. L.; Simeonov, A.; Yasgar, A.; Zheng, W.; Austin, C. P. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11473–11478.
- [9] Walters, W. P.; Namchuk, M. Designing screens: How to make your hits a hit. *Nat. Rev. Drug Discovery* **2003**, *2*, 259–266.
- [10] Copeland, R. A. Mechanistic considerations in high-throughput screening. *Anal. Biochem.* **2003**, *320*, 1–12.

- [11] Rohrer, S. G.; Baumann, K. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.* **2009**, *49*, 169–184.
- [12] Xu, H.; Agrafiotis, D. K. Retrospect and prospect of virtual screening in drug discovery. *Curr. Top. Med. Chem.* **2002**, *2*, 1305–1320.
- [13] Kossner, M. Dissertation. *Technische Universität: Braunschweig* **2011**.
- [14] Johnston, P. A.; Johnston, P. A. Cellular platforms for HTS: Three case studies. *Drug Discovery Today* **2002**, *7*, 353–363.
- [15] Mc Govern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K. A specific mechanism of nonspecific inhibition. *J. Med. Chem.* **2003**, *46*, 4265–4272.
- [16] Mc Govern, S. L.; Shoichet, B. K. Kinase inhibitors: Not just for kinases anymore. *J. Med. Chem.* **2003**, *46*, 1478–1483.
- [17] Rishton, G. M. Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today* **2003**, *8*, 86–96.
- [18] Rishton, G. M. Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today* **1997**, *2*, 382–384.
- [19] Ludewig, S.; Kossner, M.; Schiller, M.; Baumann, K.; Schirmeister, T. Enzyme kinetics and hit validation in fluorimetric protease assays. *Curr. Top. Med. Chem.* **2010**, *10*, 368–382.
- [20] Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.
- [21] Kubinyi, H. Similarity and dissimilarity: A medicinal chemist's view. *Perspect. Drug Discovery Des.* **1998**, *9-11*, 225–252.
- [22] Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350–4358.
- [23] Bender, A.; Glen, R. C. Molecular similarity: A key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
- [24] Johnson, M. A.; Maggiora, G. M. *Concepts and applications of molecular similarity*, 1st ed.; John Wiley and Sons Inc., 1990.

- [25] Maggiora, G. M. On outliers and activity cliffs - why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535.
- [26] Guha, R.; van Drie, J. Structure - activity landscape index: Identifying and quantifying activity cliffs. *J. Chem. Inf. Model.* **2008**, *48*, 646–658.
- [27] Lahana, R. How many leads from HTS? *Drug Discovery Today* **1999**, *4*, 447–448.
- [28] Ramesha, C. S. How many leads from HTS? - Comment. *Drug Discovery Today* **2000**, *5*, 43–44.
- [29] Kraemer, O.; Hazemann, I.; Podjarny, A. D.; Klebe, G. Virtual screening for inhibitors of human aldose reductase. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 814–823.
- [30] Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening - an overview. *Drug Discovery Today* **1998**, *3*, 160–178.
- [31] Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249.
- [32] Agrafiotis, D. K.; Lobanov, V. S.; Salemme, F. R. Combinatorial informatics in the post-genomics era. *Nat. Rev. Drug Discovery* **2002**, *1*, 337–346.
- [33] Kubinyi, H. Success stories of computer-aided design. In *Computer applications in pharmaceutical research and development*; Ekins, S., Wang, B., Eds., 1st ed.; Wiley Interscience, 2006; Chapter 16, pp 377–424.
- [34] Davies, J. W.; Glick, M.; Jenkins, J. L. Streamlining lead discovery by aligning in silico and high-throughput screening. *Curr. Opin. Chem. Biol.* **2006**, *10*, 343–351.
- [35] Klebe, G. Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.
- [36] Birch, P. J.; Dekker, L. V.; James, I. F.; Southan, A.; Cronk, D. Strategies to identify ion channel modulators: Current and novel approaches to target neuropathic pain. *Drug Discovery Today* **2004**, *9*, 410–418.
- [37] Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**, *3*, 928–944.

- [38] Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*, 1st ed.; Wiley-VCH, 2001.
- [39] Bender, A.; Glen, R. C. A discussion of measures of enrichment in virtual screening: Comparing the information content of descriptors with increasing levels of sophistication. *J. Chem. Inf. Model.* **2005**, *45*, 1369–1375.
- [40] Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. Comparison of correlation vector methods for ligand-based similarity searching. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 687–698.
- [41] MDL MACCS keys, *Symyx Technologies*. <http://www.accelrys.com>.
- [42] Stiefl, N.; Baumann, K. Mapping property distributions of molecular surfaces: Algorithm and evaluation of a novel 3D quantitative structure-activity-relationship technique. *J. Med. Chem.* **2003**, *46*, 1390–1407.
- [43] Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. Molecular surface point environments for virtual screening and the elucidation of binding patterns (MOLPRINT 3D). *J. Med. Chem.* **2004**, *47*, 6569–6583.
- [44] Renner, S.; Noeske, T.; Parsons, C. G.; Schneider, P.; Weil, T.; Schneider, G. New allosteric modulators of metabotropic glutamate receptor 5 (mGluR5) found by ligand-based virtual screening. *ChemBioChem* **2005**, *6*, 620–625.
- [45] Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): Theory and application. *J. Chem. Inf. Model.* **2007**, *47*, 279–294.
- [46] Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity searching in files of three-dimensional chemical structures: Analysis of the BIOSTER database using two-dimensional fingerprints and molecular field descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 295–307.
- [47] Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M. Bridging chemical and biological space: 'Target fishing' using 2D and 3D molecular descriptors. *J. Med. Chem.* **2006**, *49*, 6802–6810.
- [48] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

- [49] Dixon, J. S. Evaluation of the CASP2 docking section. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 198–204.
- [50] Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; la Londe, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- [51] Tirado-Rives, J.; Jorgensen, W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *J. Med. Chem.* **2006**, *49*, 5880–5884.
- [52] Boehm, H.-J.; Flohr, A.; Stahl, M. Scaffold hopping. *Drug Discovery Today: Technol.* **2004**, *1*, 217–224.
- [53] Renner, S.; Schneider, G. Scaffold-hopping potential of ligand-based similarity concepts. *ChemMedChem* **2006**, *1*, 181–185.
- [54] Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. 'Scaffold-Hopping' by topological pharmacophore search: A contribution to virtual screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894–2896.
- [55] Bischoff, E. Potency, selectivity, and consequences of nonselectivity of PDE inhibition. *Int. J. Impotence Res.* **2004**, *16*, S11–S14.
- [56] Sung, B.-J. et al. Structure of the catalytic domain of human phosphodiesterase 5 with bound drug molecules. *Nature* **2003**, *425*, 98–102.
- [57] Supuran, C. T.; Mastrolorenzo, A.; Barbaro, G.; Scozzafava, A. Phosphodiesterase 5 inhibitors - drug design and differentiation based on selectivity, pharmacokinetic and efficacy profiles. *Curr. Pharm. Des.* **2006**, *12*, 3459–3465.
- [58] Mehrotra, N.; Gupta, M.; Kovar, A.; Meibohm, B. The role of pharmacokinetics and pharmacodynamics in phosphodiesterase-5 inhibitor therapy. *Int. J. Impotence Res.* **2007**, *19*, 253–264.
- [59] Ehrlich, P. Über den jetzigen Stand der Chemotherapie. *Ber. Dtsch. Chem. Ges.* **1909**, 17–47.
- [60] Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, *70*, 1129–1143.

- [61] Steinhilber, D.; Schubert-Zsilavecz, R.; Roth, H. J. *Medizinische Chemie - Targets und Arzneistoffe*, 1st ed.; Deutscher Apotheker Verlag, 2005.
- [62] Molecular Operating Environment (MOE), Version 2010.10, *Chemical Computing Group Inc.* <http://www.chemcomp.com>.
- [63] Phase, Version 3.0, *Schrödinger*. <http://www.schrodinger.com/>.
- [64] Flower, R. J. The development of COX2 inhibitors. *Nat. Rev. Drug Discovery* **2003**, *2*, 179–191.
- [65] Renner, S.; Fechner, U.; Schneider, G. Alignment-free pharmacophore patterns - a correlation-vector approach. In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R. D., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 3, pp 49–79.
- [66] Rarey, M.; Dixon, J. S. Feature trees: A new molecular similarity measure based on tree matching. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471–490.
- [67] Rarey, M.; Fricker, P.; Hindle, S.; Metz, G.; Rummey, C.; Zimmermann, M. Feature trees: Theory and applications from large-scale virtual screening to data analysis. In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R. D., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 4, pp 81–116.
- [68] Brown, N.; Jacoby, E. On scaffolds and hopping in medicinal chemistry. *Mini-Rev. Med. Chem.* **2006**, *6*, 1217–1229.
- [69] Kaminski, J. J.; Rane, D. F.; Snow, M. E.; Weber, L.; Rothofsky, M. L.; Anderson, S. D.; Lin, S. L. Identification of novel farnesyl protein transferase inhibitors using three-dimensional database searching methods. *J. Med. Chem.* **1997**, *40*, 4103–4112.
- [70] de Lucca, G. V.; Lam, P. Y. S. De novo design, discovery and development of cyclic urea HIV protease inhibitors. *Drugs Future* **1998**, *23*, 987–994.
- [71] de Esch, I. J. P.; Mills, J. E. J.; Perkins, T. D. J.; Romeo, G.; Hoffmann, M.; Wieland, K.; Leurs, R.; Menge, W. M. P. B.; Nederkoorn, P. H. J.; Dean, P. M.; Timmerman, H. Development of a pharmacophore model for histamine H3 receptor antagonists, using the newly developed molecular modeling program SLATE. *J. Med. Chem.* **2001**, *44*, 1666–1674.

- [72] Barreca, M. L.; Gitto, R.; Quartarone, S.; de Luca, L.; de Sarro, G.; Chimirri, A. Pharmacophore modeling as an efficient tool in the discovery of novel noncompetitive AMPA receptor antagonists. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 651–655.
- [73] Langer, T.; Krovat, E. Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 370–376.
- [74] Hessler, G.; Baringhaus, K.-H. The scaffold hopping potential of pharmacophores. *Drug Discovery Today: Technol.* **2010**, *7*, 263–269.
- [75] Poptodorov, K.; Luu, T.; Hoffmann, R. D. Pharmacophore model generation software tools. In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R. D., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 2, pp 17–47.
- [76] Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* **2010**, *53*, 539–558.
- [77] van Drie, J. History of 3D pharmacophore searching: Commercial, academic and open-source tools. *Drug Discovery Today: Technol.* **2010**, *7*, 255–262.
- [78] Manetti, F.; Botta, M.; Tafi, A. Application of pharmacophore models in medicinal chemistry. In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R. D., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 12, pp 253–282.
- [79] Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D pharmacophore-based virtual screening. *Drug Discovery Today: Technol.* **2010**, *7*, 221–228.
- [80] Anghelescu, A. V.; de Lisle, R. K.; Lowrie, J. F.; Klon, A. E.; Xiaoming, X.; Diller, D. J. Technique for generating three-dimensional alignments of multiple ligands from one-dimensional alignments. *J. Chem. Inf. Model.* **2008**, *48*, 1041–1054.
- [81] Feng, J.; Sanil, A.; Young, S. S. PharmID: Pharmacophore identification using gibbs sampling. *J. Chem. Inf. Model.* **2006**, *46*, 1352–1359.
- [82] Podolyan, Y.; Karypis, G. Common pharmacophore identification using frequent clique detection algorithm. *J. Chem. Inf. Model.* **2009**, *49*, 13–21.

- [83] Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Deterministic pharmacophore detection via multiple flexible alignment of drug-like molecules. *J. Comput. Biol.* **2008**, *15*, 737–754.
- [84] de Todorov, N. P.; Alberts, I. L.; Esch, I. J. P.; Dean, P. M. QUASI: A novel method for simultaneous superposition of multiple flexible ligands and virtual screening using partial similarity. *J. Chem. Inf. Model.* **2007**, *47*, 1007–1020.
- [85] Zhu, F.; Agrafiotis, D. K. Recursive distance partitioning algorithm for common pharmacophore identification. *J. Chem. Inf. Model.* **2007**, *47*, 1619–1625.
- [86] Cho, S. J.; Sun, Y. FLAME: A program to flexibly align molecules. *J. Chem. Inf. Model.* **2006**, *46*, 298–306.
- [87] Marialke, J.; Koerner, R.; Tietze, S.; Apostolakis, J. Graph-based molecular alignment (GMA). *J. Chem. Inf. Model.* **2007**, *47*, 591–601.
- [88] Ligandscout, Version 3.0, *Inte:Ligand*. <http://www.inteligand.com>.
- [89] Ehrlich, P. Chemotherapeutics: Scientific principles, methods and results. *Lancet* **1913**, *2*, 445–451.
- [90] Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985–2993.
- [91] Triballeau, N.; Bertrand, H.-O.; Acher, F. Are you sure you have a good model? In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R. D., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 15, pp 325–364.
- [92] Discovery Studio, Version 3.1, *Accelrys*. <http://www.accelrys.com/>.
- [93] *Protein-ligand interactions: From molecular recognition to drug design*, 1st ed.; Boehm, H.-J., Schneider, G., Mannhold, R., Kubinyi, H., Eds.; Wiley-VCH, 2003; Vol. 19.
- [94] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.
- [95] Boehm, H.-J.; Klebe, G. What can we learn from molecular recognition in protein–ligand complexes for the design of new drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588–2614.

- [96] Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **1976**, *105*, 1–12.
- [97] Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* **1974**, *248*, 338–339.
- [98] Chothia, C.; Janin, J. Principles of protein protein recognition. *Nature* **1975**, *256*, 705–708.
- [99] Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 3086–3090.
- [100] Eisenberg, D.; Mc Lachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199–203.
- [101] Pace, C. N. Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* **1992**, *226*, 29–35.
- [102] Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P. A. Solvation model based on weighted solvent accessible surface area. *J. Phys. Chem. B* **2001**, *105*, 5055–5067.
- [103] Searle, M. S.; Williams, D. H.; Gerhard, U. Partitioning of free energy contributions in the estimation of binding constants: Residual motions and consequences for amide-amide hydrogen bond strengths. *J. Am. Chem. Soc.* **1992**, *114*, 10697–10704.
- [104] Nicholls, A.; Sharp, K. A.; Honig, B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 281–296.
- [105] Ross, P. D.; Rekharsky, M. V. Thermodynamics of hydrogen bond and hydrophobic interactions in cyclodextrin complexes. *Biophys. J.* **1996**, *71*, 2144–2154.
- [106] Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist’s guide to molecular interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.
- [107] Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Bichard, C. J.; Orchard, M. G.; Fleet, G. W.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M.; Papageorgiou, A. Design of inhibitors of glycogen phosphorylase: A

- study of alpha- and beta-C-glucosides and 1-thio-beta-D-glucose compounds. *Biochemistry* **1994**, *33*, 5745–5758.
- [108] Sauter, N. K.; Bednarski, M. D.; Wurzburg, B. A.; Hanson, J. E.; Whitesides, G. M.; Skehel, J. J.; Wiley, D. C. Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: A 500-MHz proton nuclear magnetic resonance study. *Biochemistry* **1989**, *28*, 8388–8396.
- [109] Janes, W.; Schulz, G. E. The binding of the retro-analogue of glutathione disulfide to glutathione reductase. *J. Biol. Chem.* **1990**, *265*, 10443–10445.
- [110] Hunter, C. A. Quantifying intermolecular interactions: Guidelines for the molecular recognition toolbox. *Angew. Chem., Int. Ed.* **2004**, *43*, 5310–5324.
- [111] Mills, J. E. J. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 607–622.
- [112] Bruno, I. J.; Cole, J. C.; Lommerse, J. P. M.; Rowland, R. S.; Taylor, R.; Verdonk, M. L. IsoStar: A library of information about nonbonded interactions. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525–537.
- [113] Boer, D. R.; Kroon, J.; Cole, J. C.; Smith, B.; Verdonk, M. L. SuperStar: Comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein-ligand interactions. *J. Mol. Biol.* **2001**, *312*, 275–287.
- [114] Hendlich, M. Databases for protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 1178–1182.
- [115] Hendlich, M.; Bergner, A.; Guenther, J.; Klebe, G. Relibase: Design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.
- [116] Schmitt, S.; Kuhn, D.; Klebe, G. A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.* **2002**, *323*, 387–406.
- [117] Burley, S. K.; Petsko, G. A. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. *Science* **1985**, *229*, 23–28.
- [118] Dougherty, D. A. Cation-pi interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **1996**, *271*, 163–168.

- [119] Mortimer, C. E.; Müller, U. *Chemie. Das Basiswissen der Chemie. Mit Übungsaufgaben.*, 6th ed.; Thieme, 1996.
- [120] Cole, J. C.; Taylor, R.; Verdonk, M. L. Directional preferences of intermolecular contacts to hydrophobic groups. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 1183–1193.
- [121] Gohlke, H.; Klebe, G. Drugscore meets CoMFA: Adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* **2002**, *45*, 4153–4170.
- [122] Toth, G.; Bowers, S. G.; Truong, A. P.; Probst, G. The role and significance of unconventional hydrogen bonds in small molecule recognition by biological receptors of pharmaceutical relevance. *Curr. Pharm. Des.* **2007**, *13*, 3476–3493.
- [123] Schmitt, J. D. Exploring the nature of molecular recognition in nicotinic acetylcholine receptors. *Curr. Med. Chem.* **2000**, *7*, 749–800.
- [124] Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van der Oost, J.; Smit, A. B.; Sixma, T. K. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **2001**, *411*, 269–276.
- [125] Gallivan, J. P.; Dougherty, D. A. A computational study of cation- π interactions vs salt bridges in aqueous media: Implications for protein engineering. *J. Am. Chem. Soc.* **2000**, *122*, 870–874.
- [126] Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: A knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **1999**, *289*, 1093–1108.
- [127] Gohlke, H.; Hendlich, M.; Klebe, G. Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. *Perspect. Drug Discovery Des.* **2000**, *20*, 115–144.
- [128] Bron, C.; Kerbosch, J. Finding all cliques of an undirected graph. *Commun. ACM* **1973**, *16*, 575–577.
- [129] Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- [130] Boobbyer, D. N. A.; Goodford, P. J.; Mc Whinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable

- binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.
- [131] Wade, R. C.; Clark, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140–147.
- [132] Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148–156.
- [133] Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. FLAP: GRID molecular interaction fields in virtual screening. Validation using the DUD data set. *J. Chem. Inf. Model.* **2010**, *50*, 1442–1450.
- [134] Bergmann, R.; Linusson, A.; Zamora, I. SHOP: Scaffold hopping by GRID-based similarity searches. *J. Med. Chem.* **2007**, *50*, 2708–2717.
- [135] Pastor, M.; Cruciani, G.; Mc Lay, I.; Pickett, S.; Clementi, S. GRID-independent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.
- [136] Cruciani, G.; Pastor, M.; Guba, W. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11*, 29–39.
- [137] Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204–2216.
- [138] Cruciani, G.; Carosati, E.; Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: Understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* **2005**, *48*, 6970–6979.
- [139] Milletti, F.; Storchi, L.; Sforna, G.; Cruciani, G. New and original pKa prediction method using GRID molecular interaction fields. *J. Chem. Inf. Model.* **2007**, *47*, 2172–2181.
- [140] Milletti, F.; Storchi, L.; Sforna, G.; Cross, S.; Cruciani, G. Tautomer enumeration and stability prediction for virtual screening on large chemical databases. *J. Chem. Inf. Model.* **2009**, *49*, 68–75.

- [141] Cross, S.; Cruciani, G. Molecular fields in drug discovery: Getting old or reaching maturity? *Drug Discovery Today* **2010**, *15*, 23–32.
- [142] *Molecular interaction fields: Applications in drug discovery and ADME prediction*; Cruciani, G., Ed.; Wiley-VCH, 2006; Vol. 27.
- [143] Cross, S.; Cruciani, G. GRID-derived structure-based 3D pharmacophores and their performance compared to docking. *Drug Discovery Today: Technol.* **2010**, *7*, 213–219.
- [144] Wang, R.; Liu, L.; Lai, L.; Tang, Y. SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *J. Mol. Model.* **1998**, *4*, 379–394.
- [145] Miranker, A.; Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29–34.
- [146] Caffisch, A.; Miranker, A.; Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* **1993**, *36*, 2142–2167.
- [147] Brooks, B. R. et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- [148] Dennis, S.; Kortvelyesi, T.; Vajda, S. Computational mapping identifies the binding sites of organic solvents on proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 4290–4295.
- [149] Landon, M. R.; Lancia Jr., D. R.; Yu, J.; Thiel, S. C.; Vajda, S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J. Med. Chem.* **2007**, *50*, 1231–1240.
- [150] Schubert, C. R.; Stultz, C. M. The multi-copy simultaneous search methodology: A fundamental tool for structure-based drug design. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 475–489.
- [151] Brenk, R.; Klebe, G. 'Hot Spot' analysis of protein-binding sites as a prerequisite for structure-based virtual screening and lead optimization. In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R. D., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 8, pp 171–192.

- [152] Mason, J. S.; Cheney, D. L. Ligand-receptor 3-D similarity studies using multiple 4-point pharmacophores. *Pac. Symp. Biocomput.* **1999**, *4*, 456–467.
- [153] Mason, J. S.; Cheney, D. L. Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* **2000**, *5*, 576–587.
- [154] Perruccio, F.; Mason, J. S.; Sciabola, S.; Baroni, M. FLAP: 4-Point pharmacophore fingerprints from GRID. In *Molecular interaction fields: Applications in drug discovery and ADME prediction*; Cruciani, G., Ed., 1st ed.; Wiley-VCH, 2006; Vol. 27, Chapter 4, pp 83–102.
- [155] Fox, T.; Haaksma, E. E. J. Computer based screening of compound databases: 1. Preselection of benzamidine-based thrombin inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 411–425.
- [156] Tintori, C.; Corradi, V.; Magnani, M.; Manetti, F.; Botta, M. Targets looking for drugs: A multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *J. Chem. Inf. Model.* **2008**, *48*, 2166–2179.
- [157] Asinex, *Asinex-Gold-Database*, http://www.asinex.com/Libraries_Gold_Platinum.html, zuletzt abgerufen am 14.05.2012.
- [158] von Itzstein, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418–423.
- [159] Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–169.
- [160] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [161] Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B. The protein data bank: A computer based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- [162] Wolber, G.; Dornhofer, A. A.; Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 773–788.

- [163] Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- [164] Salam, N. K.; Nuti, R.; Sherman, W. Novel method for generating structure-based pharmacophores using energetic analysis. *J. Chem. Inf. Model.* **2009**, *49*, 2356–2368.
- [165] Ortuso, F.; Langer, T.; Alcaro, S. GBPM: GRID-based pharmacophore model: Concept and application studies to protein-protein recognition. *Bioinformatics* **2006**, *22*, 1449–1455.
- [166] Ortuso, F.; Alcaro, S.; Langer, T. GRID-based pharmacophore models: Concept and application examples. In *Pharmacophores and pharmacophore searches*; Langer, T., Hoffmann, R., Eds., 1st ed.; Wiley-VCH, 2006; Vol. 32, Chapter 7, pp 151–170.
- [167] Chen, J.; Lai, L. Pocket v.2: Further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.* **2006**, *46*, 2684–2691.
- [168] Henrich, S.; Salo-Ahen, O. M. H.; Huang, B.; Rippmann, F.; Cruciani, G.; Wade, R. C. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **2010**, *23*, 209–219.
- [169] Pocket Finder, *University of Leeds*. <http://www.modelling.leeds.ac.uk/pocketfinder/>, zuletzt abgerufen am 09.05.2012.
- [170] Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Modell.* **1997**, *15*, 359–363.
- [171] Q-Site Finder, *University of Leeds*. <http://www.modelling.leeds.ac.uk/qsitefinder/>, zuletzt abgerufen am 09.05.2012.
- [172] Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908–1916.
- [173] Boehm, H.-J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.

- [174] Boehm, H.-J. A novel computational tool for automated structure-based drug design. *J. Mol. Recognit.* **1993**, *6*, 131–137.
- [175] Allen, F. H.; Kennard, O.; Taylor, R. Systematic analysis of structural data as a research technique in organic chemistry. *Acc. Chem. Res.* **1983**, *16*, 146–153.
- [176] Boehm, H.-J. LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593–606.
- [177] Grant, J. A.; Pickup, B. T. A Gaussian description of molecular shape. *J. Phys. Chem.* **1995**, *99*, 3503–3510.
- [178] Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1666.
- [179] Livingstone, D. J.; Clark, T.; Ford, M. G.; Hudson, B. D.; Whitley, D. C. QSAR studies using the parashift system. *SAR QSAR Environ. Res.* **2008**, *19*, 285–302.
- [180] Ritchie, D. W.; Kemp, G. J. L. Protein docking using spherical polar Fourier correlations. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 178–194.
- [181] Ballester, P. J.; Richards, W. G. Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* **2007**, *28*, 1711–1723.
- [182] Haque, I. S.; Pande, V. S. PAPER - accelerating parallel evaluations of ROCS. *J. Comput. Chem.* **2010**, *31*, 117–132.
- [183] Piramid, Version 1.0.1, *Silicos NV*. <http://www.silicos.be>.
- [184] Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual screening using molecular fields. Application to the DUD data set. *J. Chem. Inf. Model.* **2008**, *48*, 2108–2117.
- [185] Vainio, M. J.; Puranen, J. S.; Johnson, M. S. ShaEP: Molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* **2009**, *49*, 492–502.
- [186] Nicholls, A.; Grant, J. Molecular shape and electrostatics in the encoding of relevant chemical information. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 661–686.
- [187] de Taminiau, J.; Thijs, G.; de Winter, H. Pharao: Pharmacophore alignment and optimization. *J. Mol. Graphics Modell.* **2008**, *27*, 161–169.

- [188] Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical function queries for 3D database search. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297–1308.
- [189] FRED, Version 2.2.5, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [190] ROCS, Version 3.1, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [191] Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How to optimize shape-based virtual screening: Choosing the right query and including chemical information. *J. Chem. Inf. Model.* **2009**, *49*, 678–692.
- [192] Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.
- [193] Sutherland, J. J.; Nandigam, R. K.; Erickson, J. A.; Vieth, M. Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. *J. Chem. Inf. Model.* **2007**, *47*, 2293–2302.
- [194] Mc Gaughey, G. B.; Sheridan, R.; Bayly, C. I.; Culberson, J. C.; Kretsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519.
- [195] Rush III, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.
- [196] Bostroem, J.; Berggren, K.; Elebring, T.; Greasley, P. J.; Wilstermann, M. Scaffold hopping, synthesis and structure-activity relationships of 5,6-diarylpyrazine-2-amide derivatives: A novel series of CB1 receptor antagonists. *Bioorg. Med. Chem.* **2007**, *15*, 4077–4084.
- [197] Freitas, R. F.; Oprea, T. I.; Montanari, C. A. 2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L. *Bioorg. Med. Chem.* **2008**, *16*, 838–853.
- [198] Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative performance assessment of the conformational model generators Omega and Catalyst: A large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* **2006**, *46*, 1848–1861.

- [199] OMEGA2, Version 2.3.2, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [200] EON, Version 2.0.1, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [201] shape-it, Version 1.0.0, *Silicos NV*. <http://silicos-it.com/tools/shape-it/shape-it.html>.
- [202] de Winter, H. pers. Kommunikation, Silicos NV, 2012.
- [203] Pharaos, Version 3.0.3, *Silicos NV*. <http://www.silicos.be>.
- [204] align-it, Version 1.0.1, *Silicos NV*. <http://silicos-it.com/tools/align-it/align-it.html>.
- [205] Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: Overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251–3264.
- [206] DAYLIGHT SMARTS, *Daylight Chemical Information Service Inc.* <http://www.daylight.com>.
- [207] Mason, J. S.; Cheney, D. L. Ligand-receptor 3-D similarity studies using multiple 4-point pharmacophores. *Pac. Symp. Biocomput.* **1999**, *4*, 456–467.
- [208] Mason, J. S.; Cheney, D. L. Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* **2000**, *5*, 576–587.
- [209] Sciabola, S.; Stanton, R. V.; Mills, J. E.; Flocco, M. M.; Baroni, M.; Cruciani, G.; Perruccio, F.; Mason, J. S. High-throughput virtual screening of proteins using GRID molecular interaction fields. *J. Chem. Inf. Model.* **2010**, *50*, 155–169.
- [210] Good, A. C.; Mason, J. S. Three-dimensional structure database searches. *Rev. Comput. Chem.* **1996**, *7*, 67–117.
- [211] Good, A. C.; Cheney, D. L. Analysis and optimization of structure-based virtual screening protocols: 1. Exploration of ligand conformational sampling techniques. *J. Mol. Graphics Modell.* **2003**, *22*, 23–30.

- [212] Bostroem, J. Reproducing the conformations of protein-bound ligands: A critical evaluation of several popular conformational searching tools. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1137–1152.
- [213] Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technol.* **2010**, *7*, 245–253.
- [214] Langer, T.; Wolber, G. Pharmacophore definition and 3D searches. *Drug Discovery Today: Technologies* **2004**, *1*, 203–207.
- [215] Spitzer, G. M.; Heiss, M.; Mangold, M.; Markt, P.; Kirchmair, J.; Wolber, G.; Liedl, K. R. One concept, three implementations of 3D pharmacophore-based virtual screening: Distinct coverage of chemical search space. *J. Chem. Inf. Model.* **2010**, *50*, 1241–1247.
- [216] O’Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 1–33.
- [217] Open Babel, Version 2.3.0, *Entwickler-Konsortium*. <http://www.openbabel.org>.
- [218] Figueras, J. Ring perception using breadth-first search. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 986–991.
- [219] Roos-Kozel, B. L.; Jorgensen, W. L. Computer-assisted mechanistic evaluation of organic reactions. 2. Perception of rings, aromaticity, and tautomers. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 101–111.
- [220] Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discovery Today* **2008**, *13*, 23–29.
- [221] Wildman, S. A.; Crippen, G. M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.
- [222] Labute, P. Protonate 3D: Assignment of ionisation states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75*, 187–205.
- [223] Dreher, J. K. F. Dissertation. *Technische Universität: Braunschweig* **2011**.
- [224] Biopython, Version 1.56, *Entwickler-Konsortium*. <http://www.biopython.org>.

- [225] Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.
- [226] Hamelryck, T.; Manderick, B. PDB file parser and structure class implemented in Python. *Bioinformatics* **2003**, *19*, 2308–2310.
- [227] Dokumentation, *Bio.PDB*, http://biopython.org/DIST/docs/cookbook/biopdb_faq.pdf, zuletzt abgerufen am 09.05.2012.
- [228] O’Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: A Python wrapper for the Open Babel cheminformatics toolkit. *Chem. Cent. J.* **2008**, *2*, 1–7.
- [229] Lee, D. T.; Wong, C. K. Worst-case analysis for region and partial region searches in multidimensional binary search trees and balanced quad trees. *Acta Inf.* **1977**, 23–29.
- [230] Boehm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243–256.
- [231] Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: An annotated database of druggable binding sites from the protein data bank. *J. Chem. Inf. Model.* **2006**, *46*, 717–727.
- [232] GRID, Version 22c, *Molecular Discovery*. <http://www.moldiscovery.com>.
- [233] Bitetti-Putzer, R.; Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Functional group placement in protein binding sites: A comparison of GRID and MCSS. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 935–960.
- [234] Kastenholtz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E. J.; Fox, T. GRID/CPCA: A new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43*, 3033–3044.
- [235] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

- [236] Loving, K.; Salam, N. K.; Sherman, W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 541–554.
- [237] Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection - what can we learn from earlier mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213–228.
- [238] Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual screening workflow development guided by the 'receiver operating characteristic' curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547.
- [239] Hanley, J. A.; Mc Neil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36.
- [240] Truchon, J.-F.; Bayly, C. I. Evaluating virtual screening methods: Good and bad metrics for the 'early recognition' problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
- [241] Sheridan, R.; Singh, S. B.; Fluder, E. M.; Kearsley, S. K. Protocols for bridging the peptide to nonpeptide gap in topological similarity searches. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1395–1406.
- [242] Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.
- [243] von Korff, M.; Freyss, J.; Sander, T. Comparison of ligand- and structure-based virtual screening on the DUD data set. *J. Chem. Inf. Model.* **2009**, *49*, 209–231.
- [244] Zhao, W.; Hevener, K. E.; White, S. W.; Lee, R. E.; Boyett, J. M. A statistical framework to evaluate virtual screening. *BMC Bioinf.* **2009**, *10*, 1–13.
- [245] Sheridan, R. P. Alternative global goodness metrics and sensitivity analysis: Heuristics to check the robustness of conclusions from studies comparing virtual screening methods. *J. Chem. Inf. Model.* **2008**, *48*, 426–433.
- [246] Swamidass, S. J.; Azencott, C.-A.; Daily, K.; Baldi, P. A CROC stronger than ROC: Measuring, visualizing and optimizing early retrieval. *Bioinformatics* **2010**, *26*, 1348–1356.

- [247] Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discovery Today* **2010**, *15*, 444–450.
- [248] van Drie, J. H. Pharmacophore discovery - lessons learned. *Curr. Pharm. Des.* **2003**, *9*, 1649–1664.
- [249] Rella, M.; Rushworth, C. A.; Guy, J. L.; Turner, A. J.; Langer, T.; Jackson, R. M. Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J. Chem. Inf. Model.* **2006**, *46*, 708–716.
- [250] Davis, A. M.; Teague, S. J. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 736–749.
- [251] Carlson, H. A.; Masukawa, K. M.; Mc Cammon, J. A. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* **1999**, *103*, 10213–10219.
- [252] Carlson, H.; Mc Cammon, J. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
- [253] Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; Mc Cammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100–2114.
- [254] Clark, R. D.; Webster-Clark, D. J. Managing bias in ROC curves. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 141–146.
- [255] Nicholls, A. What do we know and when do we know it? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239–255.
- [256] Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193–199.
- [257] Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.
- [258] Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of fingerprint-based methods for virtual

- screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177–1185.
- [259] Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
- [260] Venkatraman, V.; Perez-Nueno, V. I.; Mavridis, L.; Ritchie, D. W. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J. Chem. Inf. Model.* **2010**, *50*, 2079–2093.
- [261] Kinnings, S. L.; Jackson, R. M. LigMatch: A multiple structure-based ligand matching method for 3D virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 2056–2066.
- [262] Kalliokoski, T.; Ronkko, T.; Poso, A. FieldChopper, a new tool for automatic model generation and virtual screening based on molecular fields. *J. Chem. Inf. Model.* **2008**, *48*, 1131–1137.
- [263] Irwin, J. J.; Shoichet, B. K. ZINC - a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
- [264] Ihlenfeldt, W.-D.; Takahashi, Y.; Abe, H.; Sasaki, S. Computation and management of chemical properties in CACTVS: An extensible networked approach toward modularity and compatibility. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 109–116.
- [265] Good, A. C.; Oprea, T. I. Optimization of techniques. 3. Virtual screening enrichment studies: A help or hindrance in tool selection? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 169–178.
- [266] National Center for Biotechnology Information (NCBI), *PubMed*, <http://pubchem.ncbi.nlm.nih.gov/>, zuletzt abgerufen am 09.05.2012.
- [267] Tiikkainen, P.; Markt, P.; Wolber, G.; Kirchmair, J.; Distinto, S.; Poso, A.; Kallioniemi, O. Critical comparison of virtual screening methods against the MUV data set. *J. Chem. Inf. Model.* **2009**, *49*, 2168–2178.
- [268] Shoichet Laboratory, *DUD Release 2*, <http://dud.docking.org/r2/>, zuletzt abgerufen am 09.05.2012.
- [269] Shoichet Laboratory, *PDGFRB-Ligand, DUD Release 2*, <http://dud.docking.org/r2/>, zuletzt abgerufen am 09.05.2012.

- [270] Shoichet Laboratory, *VEGFR2-Ligand, DUD Release 2*, <http://dud.docking.org/r2/>, zuletzt abgerufen am 09.05.2012.
- [271] Mackey, M. D.; Melville, J. L. Better than random? The chemotype enrichment problem. *J. Chem. Inf. Model.* **2009**, *49*, 1154–1162.
- [272] Cheeseright, T.J.; Mackey, M.D.; Melville, J.L.; and Vinter J. G., *Field-Screen: Virtual screening using molecular fields. Application to the DUD data set - supporting information*, http://pubs.acs.org/doi/suppl/10.1021/ci800110p/suppl_file/ci800110p_si_002.zip, zuletzt abgerufen am 09.05.2012.
- [273] Good, A. C., *DUD clusters*, <http://dud.docking.org/clusters/>, zuletzt abgerufen am 09.05.2012.
- [274] Martin, Y. C. Let's not forget tautomers. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 693–704.
- [275] Oellien, F.; Cramer, J.; Beyer, C.; Ihlenfeldt, W.-D.; Selzer, P. M. The impact of tautomer forms on pharmacophore-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 2342–2354.
- [276] QUACPAC, Version 1.3.1, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [277] Bostroem, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of Omega with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449–462.
- [278] FILTER, Version 2.0.2, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [279] BABEL, Version 3.3, *OpenEye Scientific Software*. <http://www.eyesopen.com>.
- [280] Steindl, T. M.; Schuster, D.; Laggner, C.; Langer, T. Parallel screening: A novel concept in pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 2146–2157.
- [281] Steindl, T. M.; Schuster, D.; Wolber, G.; Laggner, C.; Langer, T. High-throughput structure-based pharmacophore modelling as a basis for successful parallel virtual screening. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 703–715.

- [282] Steindl, T. M.; Schuster, D.; Laggner, C.; Chuang, K.; Hoffmann, R. D.; Langer, T. Parallel screening and activity profiling with HIV protease inhibitor pharmacophore models. *J. Chem. Inf. Model.* **2007**, *47*, 563–571.
- [283] Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337–344.
- [284] Deng, Z.; Chuaqui, C.; Singh, J. Knowledge-based design of target-focused libraries using protein-ligand interaction constraints. *J. Med. Chem.* **2006**, *49*, 490–500.
- [285] Kelly, M. D.; Mancera, R. L. Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1942–1951.
- [286] Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195–207.
- [287] Perez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixido, J. APIF: A new interaction fingerprint based on atom pairs and its application to virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 1245–1260.
- [288] Chemical Computing Group Inc., *svl exchange*, <http://svl.chemcomp.com/>, zuletzt abgerufen am 09.05.2012.
- [289] Grimshaw, S. pers. Kommunikation, MOE UMG, Cambridge, 2010.
- [290] Thijs, G. pers. Kommunikation, Silicos NV, 2011.
- [291] Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, *50*, 726–741.
- [292] Knockaert, M.; Greengard, P.; Meijer, L. Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol. Sci.* **2002**, *23*, 417–425.
- [293] Hardcastle, I. R.; Golding, B. T.; Griffin, R. J. Designing inhibitors of cyclin-dependent kinases. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 325–348.
- [294] Shapiro, G. I. Cyclin-dependent kinase pathways as targets for cancer treatment. *J. Clin. Oncol.* **2006**, *24*, 1770–1783.

- [295] Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the 'gatekeeper door': Exploiting the active kinase conformation. *J. Med. Chem.* **2010**, *53*, 2681–2694.
- [296] Gray, N. S.; Wodicka, L.; Thunnissen, A.-M. W. H.; Norman, T. C.; Kwon, S.; Espinoza, F. H.; Morgan, D. O.; Barnes, G.; le Clerc, S.; Meijer, L.; Kim, S.-H.; Lockhart, D. J.; Schultz, P. G. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **1998**, *281*, 533–538.
- [297] Traxler, P.; Furet, P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol. Ther.* **1999**, *82*, 195–206.
- [298] Scarpini, E.; Scheltens, P.; Feldman, H. Treatment of Alzheimer's disease: Current status and new perspectives. *Lancet Neurol.* **2003**, *2*, 539–547.
- [299] Kryger, G.; Silman, I.; Sussman, J. L. Structure of acetylcholinesterase complexed with E2020 (Aricept): Implications for the design of new anti-Alzheimer drugs. *Structure* **1999**, *7*, 297–307.
- [300] de Keersmaecker, K.; Cools, J. Chronic myeloproliferative disorders: A tyrosine kinase tale. *Leukemia* **2006**, *20*, 200–205.
- [301] Mol, C. D.; Dougan, D. R.; Schneider, T. R.; Skene, R. J.; Kraus, M. L.; Scheibe, D. N.; Snell, G. P.; Zou, H.; Sang, B.-C.; Wilson, K. P. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* **2004**, *279*, 31655–31663.
- [302] Natesh, R.; Schwager, S. L. U.; Sturrock, E. D.; Acharya, K. R. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* **2003**, *421*, 551–554.
- [303] Salonen, L. M.; Holland, M. C.; Kaib, P. S. J.; Haap, W.; Benz, J.; Mary, J.-L.; Kuster, O.; Schweizer, W. B.; Banner, D. W.; Diederich, F. Molecular recognition at the active site of factor Xa: Cation- π interactions, stacking on planar peptide surfaces, and replacement of structural water. *Chem. Eur. J.* **2012**, *18*, 213–222.
- [304] Lee, Y.-K.; Player, M. R. Developments in factor Xa inhibitors for the treatment of thromboembolic disorders. *Med. Res. Rev.* **2009**, *31*, 202–283.
- [305] Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP

- kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **2002**, *9*, 268–272.
- [306] Schrödinger, *Glide Ligand Decoys Set*, <http://www.schrodinger.com/productpage/14/5/74/>, zuletzt abgerufen am 09.05.2012.
- [307] Whittle, M.; Gillet, V. J.; Willett, P.; Loesel, J. Analysis of data fusion methods in virtual screening: Similarity and group fusion. *J. Chem. Inf. Model.* **2006**, *46*, 2206–2219.
- [308] Sheridan, R.; Mc Gaughey, G.; Cornell, W. D. Multiple protein structures and multiple ligands: Effects on the apparent goodness of virtual screening results. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 257–265.
- [309] Favia, A. D.; Bottegoni, G.; Nobeli, I.; Bisignano, P.; Cavalli, A. SERAPhiC: A benchmark for in silico fragment-based drug design. *J. Chem. Inf. Model.* **2011**, *51*, 2882–2896.
- [310] PyMOL, Version 1.1 educational, *Schrödinger*. <http://www.pymol.org/>.
- [311] Python, Version 2.6.6, *Entwickler-Konsortium*. <http://www.python.org>.
- [312] Numpy, Version 1.3.0, *Entwickler-Konsortium*. <http://www.numpy.scipy.org/>.
- [313] hcluster, Version 0.2.0, *Eads, D.* <http://www.scipy-cluster.googlecode.com/>.
- [314] Hunter, J. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 99–104.

Index

- Ähnlichkeitsprinzip, 24
- Abl-Kinase, 162
- Acetylcholinesterase, 158
- Activity Cliff, 25
- Activity Profiling, 140
- Adenosintriphosphat, 154
- Aktive, 114
- Alignment, flexibles, 37
- Analogue Bias, 126, 174
- Angiotensin-Konversionsenzym, 162
- Anreicherung, scheinbare, 125
- Antithrombotika, 164
- Aricept[®], 158
- ASINEX-GOLD-Datenbank, 56
- Assay, Bestätigungs-, 22
- Assay, Enzym, 24
- Assay, Ganzzell-, 23
- Assay, Primär-, 22
- Assay, Proliferations-, 23
- Assay, Reporter-Gen-, 24
- Atom Count, 27
- AUC-ROC, 119
- Balanced Labeling Performance, 117
- BEDROC, 120
- Best Hit, 135
- Bindungsenthalpie, 41
- Bio.PDB-Methode, 84
- Biopython, 216
- C-Pharmakophor, 60
- Cambridge
 - Crystallographic Data Center, 49
- Cats2D, 27
- Cavbase, 49
- CavKA HYBRID, 113
- CavKA, Vergleich mit Ligandscout, 151
- Celecoxib, 26, 33
- CHARMM, 52
- Chemotype Enrichment, 130
- Cialis[®], 31
- cKIT-Kinase, 162
- Clathrat, 43
- Clique Detection, 50, 70
- cyclischen
 - Guanosin-5-Monophosphat, 181
- Cyklinabhängige-Kinase, 154
- Datensatz Vorbehandlung,
 - Ligand-Rezeptor-Komplexe, 134
- Datensatz Vorbehandlung,
 - ligcent_HTS, 133
- Datensatz Vorbehandlung,
 - Referenz-Ligand, 135
- Datensatz Vorbehandlung,
 - struct_HQS, 132
- Datensatz Vorbehandlung,
 - struct_HTS, 132
- de-novo Design, 65
- Decoys, 114
- Deskriptor, Simple, 27
- DFG-in binder, 154
- DFG-out binder, 160

- Dihydrofolat, 138
Dihydrofolat-Reduktase, 138
DIP²-Deskriptor, 90
Directory of Useful Decoys, 125
Docking, 29
Docking, Posing, 29
Docking, Scoring, 29
Donepezil, 158
Drugscore, 49
DUD-2D-Bias, 126, 180
DUD-Datensatz, 125

E-Pharmakophor, 60, 112
E-Pharmakophor, MIF-basiertes, 112
Edge-to-Fac-Interaktion, 96
Edge-to-Face-Interaktion, 47
Enthalpie, 42
Entropie, 42
EON, 68
Erektile Dysfunktion, 31
Essentials, 34
Excluded Volume, 91

Face-to-Face-Interaktion, 47, 96
Faktor Xa, 164
Feature Match, 34
Features, pharmakophore, 33
Fieldscreen-Datensatz, 128
Fieldscreen-Referenz, alternative, 147
Fingerprints, molekulare, 26
FLAP, 70

Geometrical Match, 34
Gibbs Gleichung, 41
Glide XP, 60
Glivec[®], 160
Graph, molekularer, 27
GRID, 50
GRID Based Pharmakophore Model, 60

GRID-Sonden, Konturierungsgrade, 105
GRID-Sonden, Typen, 51
Guanosin-5-Monophosphat, 181

Hückel-Regel, 80
hCluster, 215
Hinge Region, 154
Hit, 22
HIV-1 Integrase, 55
HIV-1 Reverse Transkriptase, 55
Hotspots, 48
Hydrophobic Effect, 43

Imatinib, 160
Isostar, 49

Köder-Molekül, 114
Konformerengeneration,
 directed tweak, 72
Konformerengeneration, on the fly, 72
Konformerengeneration,
 rigid-body-treatment, 72

Leitstruktur-Optimierung, 56
Levitra[®], 31
Ligand-Rezeptor-Interaktion,
 kovalente, 40
Ligand-Rezeptor-Interaktion,
 nicht-kovalente, 40
Ligand-Rezeptor-Interaktionen,
 H-Brücken-, 45
Ligand-Rezeptor-Interaktionen,
 hydrophobe, 43
Ligandeigenschaften, aromatische, 80
Ligandeigenschaften, Definitionen, 79
Ligandeigenschaften,
 H-Brücken-Akzeptor-, 82
Ligandeigenschaften,
 H-Brücken-Donor-, 82
-

- Ligandeigenschaften, hydrophobe, 80
- Ligandeigenschaften, ionische, 83
- Ligandscout, 59
- ligandzentrischer Ansatz, 66
- Ligmatch, 167
- Ligsite-Algorithmus, 63
- Lisinopril, 163
- logP, 27
- Low Energy Conformer, 68
- LUDI, 65

- MACCS-Keys, 27
- Matplotlib, 216
- Max-Fusion-Rule, 180
- Maximum Unbiased Validation, 127
- me-too Präparat, 32
- Methotrexat, 138
- MIF-Information, falsch-positive, 64, 174
- MIF-Konturierung, 50, 108
- MIF-Skalierung, 107
- MINI-Punkt, 71
- MOE Interaction Potential, 51
- MOE Multi Fragment Search, 52
- Molekül-Deskriptor, 26
- Molekül-Dynamik, 55
- Molekulares Interaktions-Feld, 50
- Mr. Brightsite, 63
- Multi Copy Simultaneous Search, 52
- Must Match, 34
- MUV-Datensatz, 127

- Neuraminidase, 56
- non-leadlike binding, 24
- Numpy, 215

- Openbabel, 216
- Overfitting, 145

- P38 Mitogen aktivierte
 Protein Kinase, 167

- Parallel Screening, 140
- Partial Match, 34
- PDGFRB, 162
- Pharao, 69
- Pharmakophor Elucidation, 36
- Pharmakophor,
 IUPAC-Definition, 33
- Pharmakophor-Ableitung,
 Apo-Strukturen, 62
- Pharmakophor-Ableitung,
 komplexbasierte Methode, 57
- Pharmakophor-Modell, Gauß-Sphären, 91
- Pharmakophor-Modell, Harte-Sphären, 91
- Pharmakophor-Modell, MOE, 91
- Pharmakophor-Modell, optimiertes, 115
- Pharmakophor-Modell, Pharao, 91
- Pharmakophor-Validierung, 114
- pharmakophore Sphären,
 überschießende Anzahl, 122
- pharmakophore Sphären, verbotene, 123
- Piramid, 69
- Piroxicam, 34
- Platelet Derived Growth Factor
 Receptor Kinase Beta, 160
- PLIDriPaS, 123, 137
- PLIF, 138
- Pocket V2.0, 61
- Pocketfinder, 63
- Protein Daten Bank, 59
- Proteinstruktur, 36
- Protonate-3D, 95
- Pseudozentrum, 50
- Pybel, 216

- Qsite Finder, 63
- Qsite-Algorithmus, 63

- Raum, chemischer, 25

- re-scoring, 137
- Relenza[®], 56
- Relibase, 49
- Retinol bindendes Protein, 43
- Rezeptor, 29
- RIE, 120
- Rivaroxaban, 165
- RMSD, 34
- ROC-Anreicherungs-Faktor, 120
- ROC-Kurve, 119
- ROCS, 68
- Rofecoxib, 26, 34

- Scaffold, 31
- Scaffold-Hopp, 31
- Schlüssel-Schloss Prinzip, 38
- Screening, fokussiertes, 25
- Screening, High Throughput, 21
- Screening-PDB, 101, 138
- Sensitivität, 116
- Sensitivität, Cluster gewichtete, 130
- ShaEP, 69
- SIFt, 137
- Sildenafil, 31, 178
- Sildenafil,
 - alternative Bindungsmodi, 179
- Similarity Property Principle, 24
- Smallest Set of Smallest Rings, 80
- Solvent Accessible Surface Area, 80
- Spezifität, 116
- Superstar, 49

- Tadalafil, 31
- Tanimoto-Koeffizient, 118
- Targets Looking for Drugs, 55
- Tautomere Zustände, 132
- The Inner Filter Effect, 24
- Thioredoxin Reduktase, 55

- Thrombin, 55
- Typ I-Inhibitor, 154
- Typ II-Inhibitor, 160

- Validierung, Gauß-Modell, 118
- Validierung, Harte-Sphären-Modell, 116
- Vardenafil, 31, 178
- Viagra[®], 31
- Virtuelles Screening, 21
- Virtuelles Screening,
 - ligandbasiertes, 26
- Virtuelles Screening,
 - strukturbasiertes, 29

- Xarelto[®], 165

- Zanamivir, 56
- ZINC-Datenbank, 125

Lebenslauf

Name: Florian Kölling

Geburtsdatum: 10.07.1980

Geburtsort : Seesen

10/2011 – heute

Technische Universität Braunschweig Institut für Medizinische und Pharmazeutische Chemie

Praktikumsleiter für das fünfte Semester: 'Arzneistoffanalytik unter besonderer Berücksichtigung der Arzneibücher'

07/2007 – 06/2012

Technische Universität Braunschweig Institut für Medizinische und Pharmazeutische Chemie

Anfertigen einer Dissertation mit dem Thema: CavKA - Cavity Knowledge Acceleration.
Entwicklung einer neuen strukturbasierten Methode zur Pharmakophor-Ableitung

(Mentor: Prof. Dr. Knut Baumann)

06/2007

Approbation als Apotheker

12/2006 – 05/2007

Pharmazeut im Praktikum

Hexal AG Holzkirchen

Methodenentwicklung Polymorphie (Infrarot-, Ramanspektroskopie, Differential Scanning Calorimetry)

05/2006 – 11/2007

Pharmazeut im Praktikum

Altstadtapotheke Göttingen

Öffentliche Apotheke

10/2001 – 02/2006

Technische Universität Braunschweig

Studium der Pharmazie

09/2000 – 07/2001

Lammetalwerkstätten Lamspringe

Zivildienst in einer Werkstatt für geistig und körperlich Behinderte

1993 – 2000 (Abitur)

Jacobson Gymnasium Seesen